5-2021

DISCOVERY OF NOVEL UBIQUITIN- AND METHYLATION-DEPENDENT INTERACTIONS USING PROTEIN DOMAIN MICROARRAYS

Jianji Chen

DISCOVERY OF NOVEL UBIQUITIN- AND METHYLATION-DEPENDENT INTERACTIONS USING PROTEIN DOMAIN MICROARRAYS

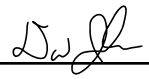
by

Jianji Chen, M.S.

APPROVED:

Mark T. Bedford, Ph.D.

Advisory Professor



David G. Johnson, Ph.D.

Taiping Chen, Ph.D.

Shawn Bratton, Ph.D.

Bin Wang, Ph.D.

APPROVED:

Dean, University of Texas

MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

DISCOVERY OF NOVEL UBIQUITIN- AND METHYLATION-DEPENDENT INTERACTIONS USING PROTEIN DOMAIN MICROARRAYS

A

Dissertation

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

by

Jianji Chen, M.S.

Houston, Texas

May, 2021

ACKNOWLEDGEMENTS

I would describe my PhD training as a winding but fruitful journey. After all the trial and error and self-reflection, I understood better what it is like to do basic science research. More importantly, the training has helped me know myself better, including what I am (not) at doing, where I want to go next, and how I can improve to be a better man.

In this section, I wanted to write something unique and memorable to thank everyone who has ever helped me.

First, I would like to thank my adviser, Dr. Mark Bedford, for all the mentorship he provided on my research, decision-making, and career development. During my PhD studies, I tend to be easily distracted by minor or irrelevant ideas that seemed exciting but were time-consuming/inconclusive. You taught me to prioritize experiments and stay focused. When I felt crushed and overwhelmed by the negative results, you would always smile and review the results meticulously to ensure that the data contained proper controls and comfort me, "It's okay. Negative data is still valuable as long as it's well controlled." You also taught me not to take experimental results personally. I need to let it go and move forward because all the evidence didn't support the original hypothesis. To give up on something that I have invested countless efforts and time in is just emotionally and psychologically challenging. However, to turn down a hypothesis is just the routine of any scientific research. A mature scientist must deal with all the negative results calmly and navigate to the alternative direction(s). You also encouraged me to participate in various collaboration projects, during which I would listen to collaborators' needs, design proper experiments, and deliver results on time. I feel accomplished by supporting collaborators to advance their projects with my expertise. Cross-functional collaboration is the core component of daily jobs for most industry positions. The more collaborations I get involved in, the more I am leaning towards a career in the industry.

I appreciate that you understand my enthusiasm for pursuing an industry position and supporting my professional development by encouraging me to do an industry internship and connecting me to your friends and former students currently working in the industry. I feel fortunate to study with you and become an emotional and scientific Grown-Up.

To all my committee faculty members, Dr. David Johnson, Dr. Shawn Bratton, Dr. Taiping Chen, Dr. Bin Wang, and Dr. Xiaobing Shi, thank you very much for attending all seven previous committee meetings and offering your resources and knowledge to advance my projects. Dr. Johnson: thank you for asking me those inspiring questions or just checking on me during my platform and poster presentation, though you might be well aware of my presentation content. I enjoyed our conversation at the Subway restaurant, and hopefully, we can have lunch together again. Dr. Bratton: Even you speak slowly, but your opinions/suggestions are always so sharp that they would nail down the cause of experimental problems. I still remember when you told me I had access to anything available in your lab. Thank you so much for your generous support. Dr. Chen is so calm that nothing seems to be able to upset him. No matter how frustrated I felt, talking with Dr. Chen would automatically "refill" my courage and confidence to stay the course and move along. Dr. Wang: thank you so much for agreeing to serve on my committee and accomodating the video conference request. I appreciate all your detailed and well-thought-out questions/feedback. Dr. Shi: thank you for giving us lectures, serving on my committee during my first year, and generously providing me with the 3xFlag-ZMYND11 vector.

I want to thank all current and previous members in Bedford lab for discussing experimental design and result interpretation and providing invaluable feedback and support along with my PhD journey: Cari Sagum: thank you for being a fantastic lab manager, ordering reagents and supplies for the entire lab, and allowing me to use three figures you generated (Figures 3, 4

and 15) in this dissertation; Dr. Guozhen Gao: thank you for getting me started on molecular biology by explaining how cloning works. Dr. Yang Fen and Dr. Yalong Wang: thank you for checking my cell culture, which allowed me to plan my experiments more strategically and use my time more efficiently. Jason Friedemann: for helping me both in the lab (protein purification and manuscript proofreading) and outside the lab (improving my English speaking and writing skills and getting me to start learning chess and financial independence/freedom). Tanner Wright: for distributing positivity to everyone by greeting with 'Happy Monday/Tuesday/Wednesday/Thursday/Friday'! Thank you for being a great team member, listener, and supporter whenever I need help. Dr. Sitaram Gayatri: thank you very much for helping me with my candidacy exam proposal, teaching me to perform the in vitro methylation assays, and prepare/test competent cells. More importantly, you taught me what information to present when proposing a new hypothesis to Dr. Bedford and other committee members. Dr. Donghang Cheng was used to working in the afternoon/evening shift. He was always there to help when I stayed up late in the evening, troubleshooting those failed pull-down experiments. He also led me to start playing tennis every Wednesday evening, which was the most exciting and enjoyable entertainment until he moved to Houston.

To all my technical mentors and collaborators in Science Park: I'd like to thank Dr. Li Wang from Dr. Sharon Dent's lab and Dr. Nicolas Veland from Dr. Taiping Chen's lab for explaining ChIP protocols step-by-step and giving me a hands-on tutorial. Dr. Liang Zhang, Dr. Rongjie Fu, and Dr. Wei He for teaching me on cloning sgRNA into CRISPR/Cas9 vector and packaging lentivirus, performing site-directed mutagenesis, and designing sgRNA, and performing bioinformatics analyses. I also would like to thank Dr. Daric Wible from Dr. Shawn Bratton's lab, Dr. Yuanz Zhong from Dr. Snow Shen's lab, Dr. Evangelis Koutelou from Dr. Sharon Dent's lab, and Dr. Tewfik Hamidi and Dr. Bigang Liu from Dr. Taiping Chen's lab, and

Dr. Yunxiang Mu from Dr. Kevin McBride's lab for sharing your reagents (antibodies and plasmids) with me and providing critical feedback on my project.

To Dr. Kevin McBride, thank you for serving in my candidacy exam committee and providing decisive guidance on custom antibody characterization and negative depletion, allowing me to move forward for the project described in **Chapter 5**.

To Dr. Collene Jeter, thank you for teaching me how to perform immunofluorescence staining, take images and conduct laser-microirradiation experiments that contributed to 3 figures (Figures 7, 12 and 19) in this dissertation. Thank you for all your efforts, from setting up the microscope to optimizing parameters.

To Dr. Maria Person and Michelle Gadush at UT Austin Proteomics Facilities: Thank you for discussing experimental design, allowing me to try different digestion enzymes, processing samples for LC-MS analysis, and delivering results and interpretation promptly (Table 6).

To Dr. Bin Liu: Thank you for helping me analyze the terabytes (TBs) of previously published proteomics data and providing a hands-on tutorial on building a bioinformatics pipeline for NGS data analysis. I appreciate your time.

To Joi Holcomb, thank you for making posters with various formats for me and designing the model figure (Figure 24) I used in this dissertation.

To Becky Brooks and Rebecca Deen, I appreciate all your help and support since Day 0, the on-site interview trip to our Science Park campus in Feb 2016. Thank you for addressing all

graduate school-related concerns and those reminder emails regarding class registration, training opportunities, and free lunch.

To all my classmates, Hieu Van, Tolkappiyan Premkumar, Yang Zeng, Dr. Martin Sarah, Melissa Frasca, and Amelie Albrecht. I would love to thank you for asking inspiring questions during my presentations and organizing off-campus happy hour events. I had so much fun with you guys at the Escape Room in downtown Austin.

To my wife, Lily, and our dog, Rex: I feel fortunate to have you right by my side through thick and thin along my PhD adventure. To my parents in China: thank you for supporting my studies in the United States. The video chat every Saturday morning always motivates me to work hard and move forward. I cannot make it this far without your companionship and unconditional support.

ABSTRACT

DISCOVERY OF NOVEL UBIQUITIN- AND METHYLATION-DEPENDENT INTERACTIONS USING PROTEIN DOMAIN MICROARRAYS

Jianji Chen, M.S.

Advisory Professor: Mark T. Bedford, Ph.D.

Post-translational modifications (PTMs) drive signal transduction by interacting with "reader" proteins. Protein domain microarray is a high throughput platform to identify novel readers for PTMs. In this dissertation, I applied two protein domain microarrays identifying novel readers for histone H2Aub1 and H2Bub1, and H3TM K4me3. Ubiquitinations of histone H2A at K119 (H2Aub1) and histone H2B at K120 (H2Bub1) function in distinct transcription regulation and DNA damage repair pathways, likely mediated by specific "reader" proteins. There are only two H2Aub1-specific readers identified and no known H2Bub1-specific readers. Using a ubiquitin-binding domain microarray, I discovered the phospholipase A2-activating protein (PLAA) PFU domain as a novel H2Bub1-specific reader. PFU domain interacts with H2Bub1 in the context of histone acid extracts but not recombinant nucleosomes, suggesting that PLAA may require additional partners for chromatin binding or PLAA only interacts with free H2Bub1. PLAA knockout cells show decreased H2Aub1 and H2Bub1, and an accumulation of a 15 kDa ubiquitin-like protein in the cytoplasm. PLAA co-localizes with laser microirradiation-induced DNA damage sites, suggesting PLAA's function in DNA damage repair. PHD fingers recognize the histone H3 N-terminal tail harboring either H3K4me3 or H3K4me0. Structural studies have identified common features among different H3K4me3 effector PHDs: Cleaved initiator methionine: a groove that fits the R2 residue, and an aromatic cage that engages the K4me3. We hypothesize that some PHDs engage with non-histone ligands whose N-termini adhere to the three rules. A search of the human proteome revealed a striking enrichment of chromatin-binding proteins, and we termed these H3 N-terminal mimicry proteins (H3TMs). We selected seven H3TMs and synthesized the methylated forms

of their N-termini. Using a methyl reader microarray, we found that they can bind known PHD and Tudor H3K4me3 effector proteins. We focused on the interaction between the kinase VRK1 and the PHF2 PHD domain. Several H3TMs peptides, in their unmethylated form, interact with NuRD complex components. These findings provide in vitro evidence that methylation of H3TMs can promote novel interactions with PHD finger- and Tudor domain-containing proteins and block interactions with the NuRD complex. We propose that these interactions can occur in vivo as well.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 PROTEIN-PROTEIN INTERACTIONS

1.1.1 Functional significance

DNA, RNA, and protein are three information-carrying biomolecules that constitute the central dogma of molecular biology (Crick, 1970). Carrying the genetic information from DNA and RNA, protein is the workhorse molecule in all living organisms. Proteins have three primary cellular functions, catalyzing chemical reactions as enzymes, transducing molecular signaling as ligand binders, and forming subcellular structures as building blocks. Execution of protein function relies on the interactions between proteins and other molecules, including protein, DNA, RNA, lipids, carbohydrates, and small molecules. Among all the intermolecular interactions, protein-protein interaction (PPI) is the best-known mechanism that orchestrates proteins' cellular functions in several different ways (Phizicky and Fields, 1995, Rao et al., 2014).

First, PPI confers substrate specificity. For example, valosin-containing protein (VCP)/p97 is an ATPase that extracts ubiquitinated target proteins in diverse cellular processes. VCP interacts with mutually exclusive ubiquitin-binding cofactors that determine the substrate specificity (Buchberger et al., 2015, Meyer et al., 2012). Second, PPI serves as an allosteric switch for enzymatic activity. A scaffold protein, Ste5, binds the mitogen-activated protein kinase Fus3 and activates Fus3 autophosphorylation by 50-fold (Bhattacharyya et al., 2006). Third, via PPI, effector proteins read the molecular signals presented by associated proteins and translate the signals into downstream biological functions. Protein post-translational modifications (PTMs) are revolutionarily conserved signals that relay the molecular messages between various proteins responding to alterations in the cellular microenvironment (Deribe et al., 2010b). Deciphering the PPI networks between histone PTMs and their effector proteins

has become a central topic in the field of epigenetics (Jenuwein and Allis, 2001a). Various histone PTMs and corresponding effector protein families have been recently reviewed (Chen et al., 2019) and will be discussed in **Chapter 3** of this dissertation. Lastly, PPI controls the formation and breakdown of subcellular structures. Monomeric actins polymerize into polarized filaments, termed F-actin, which makes up the cytoskeleton and drives cell motility (Olson and Sahai, 2009).

1.1.2 Mass spectrometry-based detection method

In the past few decades, genomic studies have identified many disease-associated gene mutations, but the underlying molecular mechanisms remain unknown. PPI plays an essential role for proteins to execute their functions. Therefore, identifying the pivotal PPI networks has proven to be a powerful approach to better understand the signaling transduction from histone codes to downstream pathways in healthy and pathological conditions.

There are two primary experimental methodologies to map the PPI network, protein domain array-based approach (discussed in **Chapter 3**) and mass spectrometry (MS)-based approaches. The MS-based approach is an unbiased discovery method with superior sensitivity. There are three ways to capture the bait protein's PPI networks, affinity purification (AP), proximity labeling (PL), and cross-linking (XL), as reviewed recently (Richards et al., 2021).

AP-MS approach uses a tagged bait protein, enabling efficient purification without immunoprecipitation-compatible antibodies (Chang, 2006, Puig et al., 2001). The bait protein can be expressed exogenously in bacteria or endogenously via ectopic overexpression. The primary limitation is the necessity of mild lysis conditions, potentially missing those transient and weak protein interactions (Rigaut et al., 1999).

Complementary to AP-MS, the PL-MS approach utilizes a bait protein fused to a non-selective labeling enzyme, such as biotin ligase, which allows biotinylation of all nearby proteins within a 10-20 nm range (Gutierrez et al., 2016). Biotinylated proteins are then enriched via streptavidin binding, making this approach better suited for detecting transient or weak protein interactions. One critique for this method is the high background signals due to non-selective labeling of all proteins nearby (Lobingier et al., 2017).

Both AP-MS and PL-MS approaches induce a tag/enzyme at the N- or C-terminus, which potentially affects the bait protein's PPI profiles and cellular functions (Sastry et al., 2009). In contrast, the XL-MS method freezes PPI networks via a chemical cross-linker such as formaldehyde, which reveals the structural information by linking the peptides near the binding interfaces without introducing artificial tags (Yu and Huang, 2018). This approach's primary limitation is that variously linked peptides pose a significant challenge to data analysis to consider all the possible combinations (Liu et al., 2015, Lu et al., 2018).

All three MS-based approaches complement each other and are proven to be powerful tools to characterize novel PPI networks in an unbiased manner.

1.2 OVERVIEW OF HISTONE H2A/H2B MONO-UBIQUITINATION

1.2.1 Protein ubiquitination

Ubiquitin is a 76-amino acid polypeptide. Ubiquitination is a process by which ubiquitin forms an isopeptide bond between its C-terminal glycine and a lysine residue from the target protein (Buetow and Huang, 2016, Komander and Rape, 2012). In addition, , Ubiquitin itself has seven lysine residues and an N-terminal methionine which can all be subjected to ubiquitination, enabling the formation of a variety of poly-ubiquitin branches/chains. A substrate can be post-

translationally modified by mono-ubiquitin or polymeric ubiquitin chains. Komander and Rape first proposed the ubiquitin code hypothesis that different ubiquitin modifications harbor unique conformations that dictate various cellular outcomes (Komander and Rape, 2012). K48 linked poly-ubiquitin chains target the modified substrates for proteasomal degradation, while K63-linked poly-ubiquitin chains and mono-ubiquitin regulate substrates' localization or interaction with other proteins.

All four core histones H2A, H2B, H3, and H4, and the linker histone H1 are heavily modified by mono- or poly-ubiquitination (Tweedie-Cullen et al., 2009). Among all the histone ubiquitination marks, mono-ubiquitination of histone H2A at K119 (H2Aub1) and H2B at K120 (H2Bub1) are relatively well characterized. H2Aub1 and H2Bub1 differ because of their unique sets of E3 ligases and deubiquitinases, distinct functions in gene transcription regulation, and DNA damage response. Notably, there are only two H2Aub1-specific effector proteins identified and no known H2Bub1-specific reader proteins (Zhang et al., 2017b, Cooper et al., 2016a).

1.2.2 Histone H2A Mono-Ubiquitination at K119

Writers and Erasers

Ubiquitination of H2A at K119 is an abundant epigenetic mark, as 10% of all nucleosomal H2As have this modification (West and Bonner, 1980, Kalb et al., 2014c, Goldknopf et al., 1975, Goldknopf and Busch, 1977). There are two major E3 ubiquitin ligases for H2Aub1, the really interesting new gene 1A/1B (RING1A/RING1B) from the polycomb repressive complex (PRC1) and the DAZ Interacting Zinc Finger Protein 3 (DZIP3) along with the N-CoR/HDAC1/3 complex (Li et al., 2006, Gray et al., 2016, Kalb et al., 2014c, Zhou et al., 2008). There are two major H2Aub1 DUBs, USP16 and breast cancer type 1 susceptibility protein (BRCA1)-associated protein 1 (BAP1) (Sahtoe et al., 2016, Gu et al., 2016).

Readers

The remodeling and spacing factor (RSF1) is a H2Aub1-specific reader protein. ChIP-Seq analyses show overlapping binding profiles of RSF1, H2Aub1, and RING1B. RSF1 knockout caused H2Aub1 chromatin organization changes and transcriptional upregulation of H2Aub1-associated genes, indicating RSF1 facilitates H2Aub1-mediated gene silencing by maintaining a stable nucleosome pattern at promoter regions (Zhang et al., 2017b).

The other H2Aub1-specific effector protein is Jarid2, a PRC2 cofactor that was first identified in Drosophila using an AP-MS approach (Cooper et al., 2016a). Jarid2 is required for PRC2 complex recruitment to the H2Aub1-modified chromatin and PRC2-mediated H3K27me3 mark (Kalb et al., 2014b, Cooper et al., 2016b). This discovery provided a mechanistic rationale for the positive-feedback-loop model between two repressive histone marks, H2Aub1, and H3K27me3, deposited by PRC1 and PCR2, respectively.

Functional characterization

Enriched at the promoters of target genes, the H2Aub1 functions as a transcriptional repressor that blocks transcription elongation by preventing FACT complex recruitment and recruiting PRC2 complex. The latter deposits the repressive H3K27me3 mark. H2Aub1 also recruits the RSF1 complex that maintains repressive chromatin conformation (Zhou et al., 2008, Cooper et al., 2016a, Kalb et al., 2014a, Cooper et al., 2014).

H2Aub1 is involved in DNA damage repair through its ability to repress local transcription at the regions flanking damage sites and also promotes homologous recombination (Kakarougkas et al., 2014).

1.2.3. Histone H2B Mono-Ubiquitination at K120

Writers and Erasers

Ubiquitination of histone H2B at K120 is present on 1% of all nucleosomal H2Bs (West and Bonner, 1980). RNF20/RNF40 complex is the major E3 ubiquitin ligase for H2Ub1, while MDM2 and BAF250/ARID1 also display H2Bub1-specific E3 ligase activity (Zhu et al., 2005, Kim et al., 2009, Minsky and Oren, 2004, Li et al., 2010). There are nine H2Bub1-targeting DUBs in human cells, USP3, USP22, USP27X, USP36, USP42, USP43, USP44, USP49, and USP51 (Nicassio et al., 2007, Zhang et al., 2008, Zhang et al., 2013, Hock et al., 2014, Gu et al., 2016, Lan et al., 2016, DeVine et al., 2018, He et al., 2018, Zhao et al., 2008, Atanassov et al., 2016). As the DUB module of the SAGA complex, USP22 is the best-characterized DUB for H2Bub1. Additionally, USP22, USP27X, and USP51 function in distinct pathways, but they compete for the same cofactors, ATXN7L3 and ENY2, to be enzymatically activated (Atanassov et al., 2016). This finding suggests a regulatory mechanism that balances the activities of different DUBs by limiting the availability of their shared adaptor proteins.

Readers

To date, there are no known H2Bub1-specific effector modules.

Functional characterization

H2Bub1 drives transcriptional activation by maintaining an open chromatin conformation, facilitating transcription elongation by RNA polymerase II together with FACT complex, promote active transcription marks, H3K4me2/me3 and H3K79me1/me2 by Set1 and DOT1L, respectively (Shiloh et al., 2011, Fierz et al., 2011, Pavri et al., 2006, Kim et al., 2009). However, the H2Bub1-specific effectors that mediate the crosstalk between H2Bub1 and those two epigenetic marks remain elusive.

H2Bub1 is required for efficient DNA damage repair by non-homologous end-joining (NHEJ) and homologous recombination (HR) repair pathways (Nakamura et al., 2011, Moyal et al., 2011). Besides, antibody class-switch recombination repair requires the removal of H2Bub1 by the SAGA complex, as H2Bub1 inhibits ATM- and DNAPK-induced γH2AX formation (Ramachandran et al., 2016, Li et al., 2018). However, H2Bub1-specific effectors for all the above functions are still yet to be determined.

1.3 PLAA IN UBIQUITIN MAINTENANCE

1.3.1 Doa1 to PLAA: Evolutionary Conservation From Yeast to Human

Melittin, a bee venom peptide, is a phospholipase A2 (PLA2) stimulatory peptide. In search of the mammalian PLA2 stimulator, Clark et al. first purified a mammalian protein using anti-melittin antibodies. They thus named the protein phospholipase A2 activating protein (PLAA) (Clark et al., 1988, Clark et al., 1987). They also found that melittin displays a high degree of sequence homology with PLAA between residues 503-538 (Bomalaski et al., 1990, Clark et al., 1991). Inflammatory mediators, including LPS, IL-1β, and TNFα, stimulated PLAA protein expression levels and increased PLA2 activity and prostaglandins (PGE2) levels. PGE2 was produced from arachidonic acid (AA) that was hydrolyzed from phospholipids by PLA2. Treatment with PLAA-specific antisense oligonucleotide caused reduced PLA2 levels (Ribardo et al., 2001). However, the molecular mechanism by which PLAA regulates PLA2 remains elusive.

PLAA's yeast ortholog, Doa1, also known as Ufd3, was first discovered as a critical player in the ubiquitin fusion degradation (UFD) pathway, together with Ufd1, Ufd2, Ufd4, and Ufd5 (Johnson et al., 1995). Both Doa1 and PLAA consist of an N-terminal WD40 domain, a central PFU domain, and a C-terminal PUL domain.

WD40 domain is a UBD and required for Doa1's functions

Pashkova et al. solved the crystal structure of the Doa1 WD40 (residues 1-300) domain with ubiquitin. The WD40 repeat β-propeller binds to mono-ubiquitin with a Kd of 220 uM, determined by chemical shift perturbation. Mutations of residues on the interaction surface disrupted ubiquitin binding and caused growth defects in yeast. Surprisingly, mutations in both WD40 and PFU domains caused synthetic growth defects (Pashkova et al., 2010).

PFU domain is a novel ubiquitin-binding module

Wilkinson and colleagues first identified Doa1 as a novel K29-linked polyubiquitin binding protein and mapped the minimal UBD residing between residue 350 and 450. They named this domain the PLAA family ubiquitin-binding (PFU) domain, as this region displays high sequence homology between Doa1 and PLAA (Mullally et al., 2006, Russell and Wilkinson, 2004). Furthermore, a Doa1-PLAA chimera protein, consisting of the Doa1 WD40 domain and PLAA PFU and PUL domains, complemented the ubiquitin depletion phenotypes and growth defects observed in Doa1-null yeast, suggesting an evolutionarily conserved function of PFU and PUL domains (Mullally et al., 2006).

The PFU domain of human PLAA is a relatively weak mono-ubiquitin binder with a Kd of 1 mM, determined by chemical shift perturbation (Fu et al., 2009a).

PUL domain interacts with Cdc48

PUL domain contains six Armadillo-repeats. Several groups demonstrated that the Doa1 PUL domain interacts with Cdc48 (Decottignies et al., 2004, Ghislain et al., 1996, Ogiso et al., 2004, Mullally et al., 2006). Structural studies showed Doa1 PUL domain binds the Cdc48 C-terminus with a Kd of 3.5 uM, determined by isothermal titration calorimetry. The PUL domain mutations that disrupt the interaction with Cdc48 caused the depletion of the cellular ubiquitin

and disruption of protein degradation, suggesting that Doa1-Cdc48 interaction is required to maintain ubiquitin homeostasis (Zhao et al., 2009).

1.3.2 Functional Characterization of PLAA/Doa1

PLAA/Doa1 is involved in four distinct pathways, ubiquitin homeostasis, DNA damage repair, lysosome damage response and mitochondrial membrane protein quality control (Johnson et al., 1995, Lis and Romesberg, 2006, Papadopoulos et al., 2017, Wu et al., 2016).

Apart from other UFD regulators (Ufd1, Ufd2, Ufd4, and Ufd5), only Doa1-null mutant showed an abnormally low content of free ubiquitin and defective degradation of Ub-P-βgal (Johnson et al., 1995). Ub-P-βgal is a model protein that is targeted and degraded by the ubiquitin fusion degradation pathway (Johnson et al., 1992). Notably, overexpression of ubiquitin rescued the defective degradation of Ub-P-βgal in Doa1-null cells, indicating Doa1 regulates degradation of Ub-P-βgal indirectly by controlling the availability of free ubiquitin.

Doa1 regulates the concentration of free ubiquitin partially by processing K48-linked polyubiquitin trimers (K48-UB3). Lis et al. found that reduced free mono-ubiquitin coincided with the accumulation of K48-UB3 but not K63-Ub3 in Doa1-null cells (Lis and Romesberg, 2006). The affinity purified Doa1 complex displayed deubiquitinase activity towards K48-linked tri-ubiquitin *in vitro*. However, it was unknown whether Doa1 itself or its associated protein(s) had the K48-Ub3 specific deubiquitinase activity. This study only tested K48-Ub3 and K63-Ub3, and it remained unknown whether Doa1 can hydrolyze other types of polyubiquitin, such as K29-Ub3, because Doa1 was originally found to bind K29-linked polyUb (Russell and Wilkinson, 2004).

Doa1 is a DNA damage "responder" by making ubiquitin available for pathways that regulate DNA replication machinery and histone H2B ubiquitination (Lis and Romesberg, 2006). Doa1-null cells displayed reduced H2Bub1. Moreover, Doa1 depletion abolished DNA damage-induced PCNA ubiquitination and H2Bub1. Interestingly, overexpression of ubiquitin only reversed the defective PCNA ubiquitination but not the H2Bub1 in Doa1-null cells, suggesting Doa1 modulates H2Bub1 via a mechanism beyond merely providing ubiquitin. This study solely focused on H2Bub1, and it was unknown how Doa1 depletion affected other histone ubiquitination marks. This result specifically interested me because I have found that PLAA is required for maintaining H2Bub1 levels in mammalian cells **in Chapter 4**.

Doa1 also functions as a ubiquitin receptor for the mitochondrial-associated degradation (MAD) pathway and it forms a complex together with Cdc48, Ufd1, and Npl4 (Wu et al., 2016). Doa1 facilitates complex recruitment by interacting with ubiquitinated proteins on the mitochondrial outer-membranes. Doa1 has two UBDs, WD40 and PFU domains. Mutations of the critical residues in WD40 and PFU domains repressed the turnover of different subsets of substrates, suggesting that these two UBDs have unique binding characteristics towards ubiquitinated substrates, the mechanisms of which have yet to be determined.

PLAA is involved in the clearance of damaged lysosomes by forming the endo-lysosomal damage response (ELDR) complex together with VCP/p97, YOD1, and UBXD1 (Papadopoulos et al., 2017). Upon localization to the ruptured lysosomes, the ELDR complex selectively removes K48-Ub conjugates but not K63-Ub conjugates. Depleting of PLAA and other two co-factors by siRNA led to the accumulation of lysosomes decorated by K48-Ub conjugates. This was consistent with another study done in yeast that discovered purified Doa1 displays a DUB activity towards K48-Ub but not K63-Ub (Lis and Romesberg, 2006).

Ubiquitin binding activity is necessary for Doa1's functions.

Depletion of Doa1 leads to decreased mono- and polyubiquitin, suggesting Doa1 is involved in ubiquitin homeostasis maintenance. Doa1-null yeast also displays increased sensitivity to protein misfolding and translation inhibition (Mullally et al., 2006). Notably, the Ub binding-deficient Doa1 mutant (F417D and F434D) only partially rescued the Doa1-null phenotypes, compared to wildtype Doa1, suggesting a functional Doa1 requires its ubiquitin-binding activity.

1.3.3 Implication of PLAA in Neurological Dysfunction

Homozygous mutations in PLAA cause severe neurological disorders in humans (Falik Zaccai et al., 2016, Hall et al., 2017). Zaccai et al. reported that PLAA L752F mutation causes progressive microcephaly and leukoencephalopathy in patients at ages 2 to 4 months (Falik Zaccai et al., 2016). They discovered that this homozygous mutation abolishes PLAA's function to induce PLA2 activity. As a result, the production of prostaglandin E2 (PGE2), a process catalyzed by PLA2, was also decreased. PLAA-null mice were perinatal lethal with reduced levels of PGE2 in the brain and lung of PLAA-null embryos. These results provided evidence that inefficient PGE2 signaling underlies the developmental defect caused by a non-functional PLAA.

Hall et al. discovered another hypomorphic PLAA mutation, G23V, that also causes severe developmental delay and seizures in affected infants (Hall et al., 2017). They found the G23V mutation impairs PLAA's function in the ubiquitin-mediated trafficking of membrane proteins for lysosomal degradation, causing reduced synaptic vesicle numbers and altered neurotransmission. PLAA-null embryos also died in mid-gestation while PLAA G23V/G23V are viable. The G23V mutation destabilized PLAA, with the mutant protein's abundance at only 30% of the wild-type counterpart. These results elucidated PLAA's role in regulating

membrane components' turnover in both central and peripheral synapses, which provided mechanistic insights on PLAA G23V mutation-caused neurological dysfunction.

1.4 HISTONE H3K4 RECOGNITION MODULES

Serving as a central hub and carrying a variety of histone PTMs, histone N-terminal tails orchestrate the recruitment of protein complexes that change the chromatin landscape at the target gene loci (Jenuwein and Allis, 2001a, Kouzarides, 2007, Ruthenburg et al., 2007b, Taverna et al., 2007, Strahl and Allis, 2000a). A combination of different PTMs on histone tails, including methylation, acetylation, and phosphorylation, has been termed the “histone code” (Jenuwein and Allis, 2001a). Single or combinatorial histone codes dictate the formation of specific chromatin-associated protein complexes that are involved in chromatin remodeling and transcriptional regulation. As an active transcription mark, histone H3K4me3 is found at the transcription start site (TSS) of almost every active gene (Bernstein et al., 2005, Ng et al., 2003, Pokholok et al., 2005, Santos-Rosa et al., 2002, Schneider et al., 2004, Schübeler et al., 2004). Transcriptional initiation encoded by H3K4me3 is accomplished by the direct recruitment of reader modules from the transcription machinery and chromatin remodeling complex (Vermeulen et al., 2007, Wysocka et al., 2006). There are two major H3K4me3 reader families: the royal superfamily (chromodomains of CHD1 and Tudor domains of JMJD2A), and the PHD-finger superfamily (Sims et al., 2005, Flanagan et al., 2005, Maurer-Stroh et al., 2003, Bienz, 2006, Huang et al., 2006a). This study is focused on the PHD-finger superfamily because of its unique ability to recognize histone H3 N-terminal tails in a modification-specific manner.

1.4.1 PHD Domain Superfamily Proteins

PHD fingers are characterized by a conserved zinc-coordinating Cys4-His-Cys3 motif (Aasland et al., 1995). On one hand, PHD fingers exhibit high sequence variability, providing various binding specificity towards H3K4me0 and K3K4me2/3 as the two major PHD ligand classes (Fiedler et al., 2008, Ramón-Maiques et al., 2007, van Ingen et al., 2008, Hung et al., 2009, Taverna et al., 2006, Wang et al., 2009a, Frottin et al., 2006, Wen et al., 2010, Ooi et al., 2007, Org et al., 2008, Chakravarty et al., 2009, Lan et al., 2007) and towards H3R2me0, H3R2me2, and H3K14ac as minor ligands (Zeng et al., 2010, Chignola et al., 2009). On the other hand, PHD fingers share two adjacent ligand-binding surfaces that stabilize the positively charged side chains of R2 and K4. These two residues determine whether the H3 N-terminal R2-T3-K4 motif can interact with PHD fingers (Ruthenburg et al., 2007a). A clear example of this can be observed in BPTF (PHD), whose R2 is anchored in place by hydrogen bonds and electrostatic interactions with negatively charged D27 while its K4me3 is stabilized by van der Waals and cation-p interactions within an aromatic cage. Notably, the free N-terminal amine forms a pair of hydrogen bonds with adjacent backbone carbonyls, enhancing the binding specificity of BPTF-H3K4me3 (Wysocka et al., 2006, Ruthenburg et al., 2011). The free N-terminus of histone H3 is generated by enzymes that facilitate the cleavage of the initial methionine (iMet) (Brandt et al., 1974, Song et al., 2003).

1.4.2 Initiator Methionine Cleavage

Initiator methionine cleavage is a co-translational process that is mediated by the methionine aminopeptidase (MAP). There are two MAPs in yeast and humans, MAP1 and MAP2 (Giglione et al., 2004, Bradshaw et al., 1998). Around two-thirds of proteins in the human proteome undergo initial methionine cleavage. The specificity of this process is determined

by the size of the P1' residue, which is the amino acid that is adjacent to the iMet (Frottin et al., 2006, Sherman et al., 1985, Wingfield, 2017b). The initiator methionine is usually cleaved when the P1' residues have side chains with a gyradius of 1.29 Angstroms (Å) or less. Seven amino acids are permissible for iMet cleavage: glycine, alanine, serine, threonine, cysteine, proline, and valine. Histone H3 has alanine (A1) at the P1' position; therefore, its iMet can be cleaved, generating a free N-terminus that stabilizes the H3-PHD finger interactions (Ruthenburg et al., 2011, Wysocka et al., 2006, Frottin et al., 2006).

1.5 HISTONE MIMICRY

Non-histone proteins with a histone-like sequence motif are called histone mimics. These proteins have the ability to mimic the H3 amino tail's interaction with modification-specific recognition modules (Marazzi et al., 2012, Sampath et al., 2007). One example of these mimics is the histone H3 methyltransferase G9a, which carries a ^{163}ARKT^{166} motif that is similar to the ^{7}ARKS^{10} motif of histone H3. Similar to H3K9, the K165 residue can be methylated by G9a, and G9a K165me2/3 interacts with the H3K9me2/3 effector HP1 (Bannister et al., 2001, Sampath et al., 2007).

Another example is the nonstructural protein 1 (NS1) of the influenza A virus H3N2 subtype. NS1 has a C-terminal ^{226}ARSK^{229} motif that resembles the N-terminal ^{1}ARTK^{4} motif of histone H3. This C-terminal motif interacts with the human PAF1 transcription elongation complex (hPAF1c), suppressing antiviral gene expression by blocking hPAF1c-mediated transcription elongation (Marazzi et al., 2012). NS1 K229me2, like dimethylated H3K4, interacts with the double chromodomains of chromatin remodeler, CHD1 (Qin et al., 2014). In addition, we have identified an internal H3-like ^{173}ARTK^{176} motif in PRMT6 that can be automethylated at the R residue (not shown). Notably, the ART/SK motifs of both NS1 and G9a are located at the C-terminal regions rather than N-terminal regions like histone H3, which suggests methylation

of the H3-mimicking motifs are critical, and location of the motifs, to a lesser extent, for hijacking chromodomains (Qin et al., 2014, Sampath et al., 2007). This also aligns with characterization of chromodomains as specialized methyl-lysine reader domains (Yap and Zhou, 2011).

In contrast, PHD fingers, the H3 N-terminal methyl-lysine reader families, have no known non-histone ligands identified. We started the project by asking whether there are proteins in the human proteome that have an H3-like N-terminal motif and whether they can interact with known H3K4me3-reader proteins in a methylation-dependent manner, which could dramatically expand the PHD interaction network.

CHAPTER 2: MATERIALS AND METHODS

Development of ubiquitin-binding domain and methyl reader domain microarrays: Protein domain microarray generation and probing have been published (Espejo and Bedford, 2004a). There are 150 ubiquitin-binding domains and 225 methyl reader domains on the two microarray chips, respectively (**Table 1** and **Table 2**).

Table 1: Proteins on the ubiquitin-binding domain microarray.

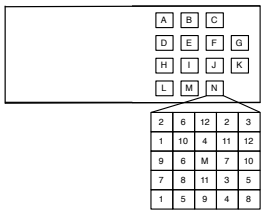
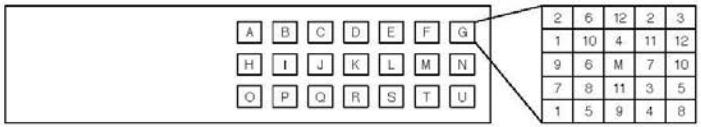
CUE **A1)** AMFR **A2)** ASCC2 **A3)** AUP1 **A4)** CUEDC1 **A5)** CUEDC2 **A6)** SMARCAD1(1-2) **A7)** TAB2 **A8)** TAB3 **A9)** TOLLIP	**JAB/MPN** **B1)** MPND **B2)** MYSM1 **B3)** PRPF8 **B4)** PSMD14 **B5)** PSMD7 **B6)** STAMBPL1	**UBA-like** **C1)** C6orf106 **C2)** Cezanne/OTUD7B **C3)** DCUN1D1 **C4)** DCUN1D2 **C5)** FAM100A/UBALD1 **C6)** FAB100B/UBALD2 **C7)** NSFL1C **C8)** OTUD7A **C9)** TTRAP/TDP2 **C10)** USP25	
UBA **D1)** LATS1 **D2)** LATS2 **D3)** MARK1 **D4)** blank **D5)** MARK3 **D6)** MARK4 **D7)** NACA **D8)** NACA2 **D9)** NBR1 **D10)** NICE4/UBAP2L **D11)** NYREN18/NUB1(1-3) **D12)** p62/SQSTM1	**UBA** **E1)** RAD23A(1-2) **E2)** RAD23B(1-2) **E3)** RHBDD3 **E4)** STS2/UBASH3A **E5)** TNRC6C **E6)** UBAC1(1-2) **E7)** UBAC2 **E8)** UBE2K **E9)** UBL7 **UBA-like+UBA** **E10)** TNK1 **UBA+UIM** **E11)** HUWE1	**UBA** **F1)** UBAP2 **F2)** UBQLN1 **F3)** UBQLN2 **F4)** UBQLN3 **F5)** UBQLN4 **F6)** UBXN1 **F7)** UBXN7 **F8)** USP13(1-2) **F9)** USP24 **F10)** USP5(1-2) **F11)** VPS13D **UBA-like+UIM** **F12)** USP25	**UBA** **G1)** CBL **G2)** CBLB **G3)** EFTS/TSFM **G4)** ETEA/FAF2 **G5)** FAF1 **G6)** TDRD3 **UBA+SH3** **G7)** STS-1/UBASH3B **UBA+Tudor** **G8)** TDRD3
UIM **H 1)** ANKIB1 **H 2)** ANKRD13/ANKRD13A(1-4) **H 3)** ANKRD13D(1-4) **H 4)** Ataxin-3L/ATXN3L(1-2) **H 5)** Ataxin-3/ATXN3(1-3) **H 6)** DNAJB2(1-2) **H 7)** EPS15(1-2) **H 8)** Epsin-1/EPN1(1-3) **H 9)** Epsin-2/EPN2(1-2) **H 10)** Epsin-3/EPN3(1-2) **H 11)** PSMD4(1-2) **H 12)** RAP80/UIMC2(1-2)	**UIM** **I1)** RNF166 **I2)** UBXN7 **I3)** USP25(1-2) **I4)** USP37(1-3) **I5)** ZFAND2B(1-2) **I6)** Hrs/HGS **I7)** blank **I8)** STAM2 **UIM+SH3** **I9)** STAM1/STAM **I10)** STAM2 **UIM+UBX** **I11)** UBXN7	**GAT** **J1)** GGA1 **J2)** GGA2 **J3)** GGA3 **J4)** Srcasm/TOM1L1 **J5)** TOM1 **J6)** TOM1L2 **VHS+GAT** **J7)** GGA1 **J8)** GGA2 **J9)** GGA3 **J10)** Srcasm/TOM1L1 **J11)** TOM1 **J12)** TOM1L2	**VHS** **K1)** GGA2 **K2)** GGA3 **K3)** Hrs/HGS **K4)** Srcasm/TOM1L1 **K5)** STAM1/STAM **K6)** STAM2 **K7)** TOM1 **K8)** TOM1L2 **VHS+UIM+SH3** **K9)** STAM1/STAM **K10)** STAM2
UMI+MIU **L1)** RNF168(UMI+MIU1) **MIU** **L2)** RNF168(2) **L3)** RNF169(2) **PFU** **L4)** PLAA **UBX** **L5)** ETEA/FAF2 **L6)** FAF1 **L7)** NSFL1C **L8)** UBXN1 **L9)** UBXN7 **UEV** **L10)** UEVLD	**Znf** **M1)** TAX1BP1(1-2) **M2)** USP13 **M3)** USP20 **M4)** USP3 **M5)** USP33 **M6)** USP39 **M7)** USP44 **M8)** USP49 **M9)** BRAP2 **ZnF+UBP** **M10)** USP5	**UBM** **N1)** hPol iota (1-2) **N2)** REV1L **UBZ** **N3)** hPol eta **N4)** hPol kappa (1-2) **N5)** PAD18 **N6)** SPARTAN **Other** **N7)** hRPN13/ADRM1 FL	

Table 2: Proteins on the methyl reader domain microarray.



TUDOR A 1) TDRD1-2 (NP_942090.1) A 2) TDRD1-3 (NP_942090.1) A 3) TDRD1-4 (NP_942090.1) A 4) TDRD2/TDRKH (NP_006053.2) A 5) TDRD3 (NP_110421.1) A 6) TDRD4-1/RNF17 (NP_112567.2) A 7) TDRD4-3/RNF17 (NP_112567.2) A 8) TDRD4-4/RNF17 (NP_112567.2) A 9) TDRD5 (NP_001105014.1) A 10) TDRD6-5 (NP_001161831.1) A 11) TDRD6-6 (NP_001161831.1) A 12) TDRD6-7 (NP_001161831.1)	TUDOR B 1) TDRD8 (NP_001247434.1) B 2) TDRD9 (NP_694561.2) B 3) TDRD10 (NP_001001945.1) B 4) TDRD11/SND1/p100TSN (NP_055205.2) B 5) TDRD12-1 (NP_001104292.1) B 6) 53BP1(1-2) (NP_005648.1) B 7) AKAP1 (NP_003479.1) B 8) ARID4A (NP_002883.3) B 9) ARID4B (NP_001193723.1) B 10) JMJD2A(1-2) (NP_055478.2) B 11) LIN9 (NP_775106.2) B 12) Pombe1 (CAA22823) *	TUDOR C 1) PHF1 (NM_002636.4) C 2) PHF19 (NM_001206840.1) C 3) PHF20 (NP_057520.2) C 4) PHF20-2 (NP_057520.2) C 5) PHF20L1 (NP_057102.4) C 6) SETDB1 (NP_001138887.1) C 7) SGF29 (NP_612423.1) C 8) SMN (NP_000335.1) C 9) SPF30/SMNDC1 (NP_005862.1) C 10) UHRF1 (NP_001041666.1) C 11) ZGPAT (NP_115916.3)	TUDOR+PHD D 1) JMJD2B/KDM4B (AA144293.1) D 2) JMJD2C/KDM4C (NP_055876.2) D 3) MTF2 (NP_001157564.1) MBT D 4) L3MBTL1(1-3) (NP_056293.4) D 5) L3MBTL3 (NP_115814.1) BAH D 6) ORC1 (NP_001177747.1) TUDOR D 7) SPIN1 FL (Q9Y657) D 8) SPIN2A FL (Q99855) D 9) SPIN2B FL (Q9BP22) D 10) SPIN3 FL (Q5JUX0) D 11) SPIN4 FL (Q56A73)	AGENET E 1) FMR1 (NP_002015.1) E 2) FXR1 (NP_005078.2) E 3) FXR2 (NP_004851.2) AGENET+KH E 4) FMRP (Q06787.1) ANK E 5) GLP (NP_079033.4) E 6) IkB FL *2 HEAT E 7) N-CAPD3(1-4) (NP_056076.1) E 8) N-CAPD3(5-8) (NP_056076.1) OTHER E 9) CYB5B FL (NP_085056.2) E 10) FKBP12 FL (AAA50476) E 11) MeCP2 **	HORMA F 1) HORMAD1 (Q86X24) F 2) HORMAD2 (Q8N7B1) F 3) MAD2L1 (Q13257) F 4) MAD2L2 (Q9UI95) F 5) hMAD2 WT (134) (NM_002358.3) ** YTH F 6) YTHDC1 (Q96MU7) F 7) YTHDC2 (Q9H6S0) F 8) YTHDF1 (Q9BYJ9) F 9) YTHDF2 (Q9Y5A9) F 10) YTHDF3 (Q7Z739) IDCL F 11) PCNA (NP_872590.1)	PWWP G 1) BRD1 (Q9588) G 2) BRPF3 (Q9ULD4) G 3) DNMT3A (Q9Y6K1) G 4) DNMT3B (Q9UBC3) G 5) HDGF (P51858) G 6) HDGF2/HDGFRP2 (Q7Z4V5) G 7) HDGFL1 (Q8TGJ6) G 8) HDGFRP3 (Q9Y3E1) G 9) LEDGF/PSIP1 (O75475) G 10) MSH6 (P52701) G 11) MUM1L1 (Q5H9M0) G 12) MBD5 (Q9P267)
CHROMO H 1) HP1β/CBX1 (NP_006790.1) H 2) CBX2 (NP_005180.1) H 3) HP1γ/CBX3 (NP_009207.2) H 4) CBX4 (NP_003646.2) H 5) HP1α/CBX5 (NP_001120793.1) H 6) CBX6 (NP_055107.2) H 7) CBX7 (NP_783640.1) H 8) CBX8 (NP_066700.1) H 9) CDY1 (NP_733941.1) H 10) CDYL1b (AAI06726.1) H 11) CDYL2 (BAC04720.1) H 12) MSL3 (NP_523353.2)	CHROMO I 1) ARID4A (NP_002883.3) I 2) ARID4B (NP_057458.4) I 3) MPP8 (NP_059990.2) I 4) MRG15 (NP_006782.1) I 5) MYST1/MOF (NP_892003.2) I 6) SMARCC1 (NP_003065.2) I 7) SMARCC2 (NP_003066.2) I 8) SUV39H1 (NP_003164.1) I 9) SUV39H2 (NP_1180353.1) I 10) TIP60 (NP_874369.1)	CHROMO J 1) CHD1 (NP_001261.2) J 2) CHD2 (NP_001262.3) J 3) CHD5 (NP_115597.3) J 4) CHD7 (NP_060250.2) J 5) CHD8 (NP_1164100.1) J 6) CHD9 (NP_079410.4) CHROM.CH+PHD J 7) CHD3 (Q12873) J 8) CHD4 (Q14839) J 9) CHD5 (Q8TDI0)	PHD K 1) ASXL1(Q8IXJ9) K 2) ASXL2 (Q76L83) K 3) ASXL3 (Q9C0F0) K 4) TAF3 (NP_114129.1) K 5) DATF1/DIDO1 (Q9BTC0) K 6) RAG2 (P55895) K 7) RSF1/HBXAP (Q96T23) K 8) SHPRH (Q149N8) K 9) TCF19 (Q9Y242) PHD(2) K 10) AIRE (O43918) K 11) MTF2 (Q9Y483)	PHD L 1) ING1 (Q9UK53) L 2) ING2 (NP_001555.1) L 3) ING3 (Q9NXR8) L 4) ING4 (Q9UNL4) L 5) ING5 (Q8WYH8) L 6) PYGO1 (Q9Y3Y4) L 7) PYGO2 (Q9BRQ0) L 8) UHRF1 (Q96T88) L 9) UHRF2 (Q96PU4) PHD(2) L 10) KAT6A/MYST3 (Q92794) L 11) KAT6B/MYST4 (Q8WYB5)	PHD M 1) KDM5A(1) (P29375) M 2) KDM5A(2) (P29375) M 3) KDM5A(3) (P29375) M 4) KDM5B(1) (Q9UGL1) M 5) KDM5B(2) (Q9UGL1) M 6) KDM5B(3) (Q9UGL1) M 7) KDM5C(1) (P41229) M 8) KDM5C(2) (P41229) M 9) KDM5D(1) (Q9BY66) M 10) KDM5D(2) (Q9BY66) M 11) KDM7A (Q6ZMT4)	PHD N 1) NSD1(4) (Q96L73) N 2) WHSC1/MMSET/NSD2(4) (Q96028) N 3) WHSC1L1/NSD3(4) (Q9BZ95) PHD(2) N 4) NSD1(1-2) (Q96L73) N 5) WHSC1/MMSET/NSD2(1-2) (Q96028) N 6) WHSC1L1/NSD3(1-2) (Q9BZ95) PHD+PWWP N 7) NSD1(3) (Q96L73) N 8) WHSC1/MMSET/NSD2(3) (Q96028) N 9) WHSC1L1/NSD3(3) (Q9BZ95)
PHD+BROMO+BAH O 1) ASH1L(long) (Q9NR48) PHD+BROMO+PWWP O 2) ZMYND9/RACK7 (Q9ULU4) O 3) ZMYND11 (Q15326) PHD+BROMO+SAND O 4) SP100 (NP_0010738601 O 5) SP110 (Q9HB58) O 6) SP140 (Q13342) O 7) SP140L (Q9H930) PHD+ODT+WHIM1 O 8) BPTF/FAL2(1) (Q12830) PHD+BROMO O 9) FAL2/BPTF(2) (Q12830)	PHD P 1) BAZ1A (Q9NRL2) P 2) TRIM28 (Q13263) PHD+BROMO P 3) BAZ1B (Q9UIG0) P 4) BAZ2A (Q9UIF9) P 5) BAZ2B (Q9UIF8) P 6) TRIM24 (O15164) P 7) TRIM28 (Q13263) P 8) TRIM33/TIF1G (Q9UPN9) P 9) TRIM66 (O15016) P 10) CBP/CREBBP (Q92793) P 11) p300/EP300 (Q09472)	PHD Q 1) PHF2 (O75151) Q 2) PHF3 (Q92576) Q 3) PHF8/JHDM1F (Q9UPP1) Q 4) PHF12(1) (Q86QT6) Q 5) PHF12(2) (Q86QT6) Q 6) PHF13 (Q86YI6) Q 7) CXXC1/PHF18 (Q9P0U4) Q 8) PHF20 (Q86V01) Q 9) PHF20L1 (A8MW92) Q 10) PHF22/INTS12 (Q96CB8) Q 11) PHF23 (Q8IUL6) Q 12) PHF24 (Q9UPV7)	PHD R 1) PHF14(3) (Q9H8G0) R 2) PHF21B (Q96EK2) PHD(2) R 3) PHF1 (O43189) R 4) PHF10 (Q8WUB8) R 5) PHF19 (Q5T6S3) PHD+ZF-C2HC R 6) PHF14(1-2) (Q9H0D0) R 7) PHF15/JADE2 (Q9NQC1) R 8) PHF16/JADE3 (Q92613) R 9) PHF17/JADE1 (Q8IE81) PHD+AT hook R 10) PHF21A (Q96BD5)	PHD+C2HC S 1) BRPF1/Peregrin (P55201) S 2) BRPF3 (Q9ULD4) S 3) KMT2A/MLL1(4) (Q03164) S 4) KMT2B/MLL2(4) (Q9UMN6) S 5) KMT2C/MLL3(1-4) (Q8NEZ4) S 6) KMT2C/MLL3(6) (Q8NEZ4) S 7) KMT2D/MLL4(1-3) (O14686) S 8) KMT2D/MLL4(7) (O14686) S 9) MLLT6 (P55196) S 10) MLLT10 (P55197)	PHD (ring) T 1) DTX4 (Q9Y2E5) PHD(2) T 2) DPF1 (Q92782) T 3) DPF3 (Q92784) PHD(2)+C2H2 T 4) DPF2 (Q92785) T 5) DPF3 (Q92784) PHD+C2H2+PHD(2) T 6) BRD1 (Q9588) PHD T 7) KMT2A/MLL1(1-3) (Q03164) T 8) KMT2B/MLL2(1-3) (Q9UMN6) T 9) KMT2C/MLL3(5-7) (Q8NEZ4) T 10) KMT2D/MLL4(4-6) (O14686) T 11) KMT2E/MLL5 (Q8IZD2)	ZF-CW U 1) KDM1B/AOF1/LSD2 (Q8NB78) U 2) MORC1 (Q86VD1) U 3) MORC2 (Q9Y6X9) U 4) MORC3 (Q14149) U 5) MORC4 (Q8TE76) ZF-CW+PWWP U 6) 2CWPW1 (Q8H0M4) U 7) 2CWPW2 (Q504Y3) ZF-ring+PHD U 8) PHRF1 (Q9P1Y6) ZF-CXXC+PHD U 9) FBXL19 (Q6PCT2)

Probe Design and Synthesis: Ubiquitination and methylation (and unmodified form) probes were synthesized and purified by LifeSensors and CPC Scientif Inc., respectively (**Table 3**). The peptides, received in powder form, were dissolved in milli-Q water and stored at -20 °C.

Table 3: Peptide probes used in this dissertation

Protein (location)	Probe Sequence Information	Vendor
Histone H2A (112-129)	Biotin-Ahx-QAVLLPKK(Ub)TESHHKAKGK	LifeSensors
Histone H2B (113-125)	Biotin-Ahx-EGTKAVTK(Ub)YTSSK	LifeSensors
VRK1 (2-13)	PRVK(me3)AAQAGRQS-Biotin	CPC Scientific Inc.
BCL11B (2-13)	SRRK(me3)QGNPQHLS-Biotin	CPC Scientific Inc.
TSHZ1 (2-13)	PRRK(me3)QQAPRRSA-Biotin	CPC Scientific Inc.
HIVEP1 (2-13)	PRTK(me3)QIHPRNLR-Biotin	CPC Scientific Inc.
PCLAF (2-13)	VRTK(me3)ADSVPGTY-Biotin	CPC Scientific Inc.
MAPK8(2-13)	SRSK(me3)RDNNFYSV-Biotin	CPC Scientific Inc.
MEF2C(2-13)	GRKK(me3)IQITRIMD-Biotin	CPC Scientific Inc.
HIRIP3(2-13)	AREK(me3)EMQEFTRS-Biotin	CPC Scientific Inc.
Histone H3 (2-19):	ARTK(me3)QTARKSTGGKAPRK-Biotin	CPC Scientific Inc.

Peptide pull-down assays: Biotinylated peptides (6 μg) were immobilized on 25 ul of streptavidin beads (Sigma, St Lous, MO, USA) in 500 μL of binding buffer (50 mM Tris-HCL pH 7.5, 150 mM NaCl, 1mM EDTA, 1mM EGTA, 0.5 % NP-40) at 4 °C for 1 hr. The beads were washed with binding buffer three times and incubated with 10 μg of glutathione S-transferase (GST) fusion protein or 30 ug 293T cell lysates for 1 hr with rotation at 4 °C. After three washes with binding buffer, the beads were boiled in 2X protein loading buffer and

subjected to SDS-PAGE followed by western blot analyses. Western blot analysis was performed as described previously (Kim et al., 2006a).

Nucleosomal pull-down assays: Three recombinant nucleosomes, unmodified dNuc, H2AK119ub dNuc, and H2BK120ub dNuc, were generously provided by Dr. Zu-Wen Sun from EpiCypher (Durham, NC). Mono-nucleosomes were assembled from recombinant human histones expressed in E. coli and wrapped by the 601 positioning sequence linked with a 5' biotin-TEG group (Lowary and Widom, 1998). Nucleosomal binding assays were performed under the same conditions as the peptide pull-down assays.

Antibodies: For western blots**:** anti-PLAA (Santa Cruz, 1:1000), anti-PLAA (Proteintech, 1:1000), anti-GFP (Santa Cruz, 1:1000), anti-H2B K120ub1 (Cell Signaling, 1:2000), anti-H2A K120ub1 (Cell Signaling, 1:2000), anti-H3 (Abcam, 1:5000), anti-flag (Sigma-Aldrich,1;5000), anti-VCP (Cell Signaling, 1:2000), anti-PHF2 (Cell Signaling, 1:2000), anti-MTA1 (Cell Signaling, 1:2000), anti-ING2 (Proteintech, 1:1000), anti-SPIN1 (ThermoFisher Scientific, 1:2000), and anti-GST (Covance, 1:5000), HRP-conjugated streptavidin antibody (ThermoFisher Scientific, 1:5000)

For immunofluorescence (IF): anti-γH2AX (Millipore, 1:200), anti-PLAA (Proteintech, 1:200), anti-H2B K120ub1 (Cell Signaling, 1:200), F-actin (Sigma-Aldrich, 1:5000), Alexa Fluor 488 goat anti-rabbit IgG, Alexa Fluor 488 goat anti-mouse IgG, Alexa Fluor 594 goat anti-rabbit IgG, and Alexa Fluor 594 goat anti-mouse IgG (ThermoFisher Scientific, 1:500)

Cell culture: MCF-7, 293T, Hela, Huh-7, Hep-G2 and Hep-3B cell lines were tested to be mycoplasma free using the MycoAlert™ kit (Lonza) and were cultured in DMEM (Gibco, 12100046) supplemented with 10% fetal bovine serum (Gibco, heat-inactivated), MEM non-

essential amino acids (Gibco, 11140050) and penicillin-streptomycin (Gibco, 50 U/mL, final concentration)

CRISPR/Cas9 plasmid construction and lentiviral packaging: 5 μg of the lentiCRISPRv2 plasmid was digested with 3 μL BsmBI (NEB, R0739) in 60 μL 1X NEBuffer^{TM} 3.1 (NEB) for 1 hour at 55 °C and the 11 kb backbone fragment was purified by 0.8% agarose gel electrophoresis and PCR & DNA Cleanup Kit (Monarch, T1030L).

A single guide RNA (sgRNA) targeting the Exon 1 of PLAA was designed using the Broad Institute GPP (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). Oligos, synthesized by Integrated DNA Technologies (IDT), were phosphorylated using the T4 PNK (NEB M0201S) in 10 μL T4 ligase buffer (NEB). The product was first denatured at 95 °C for 5 min, then slowly annealed to 25°C at 5 °C/min.
The annealed oligos and the 11 kb backbone were ligated using the Quick Ligase (NEM M2200S) and transformed into Stbl3 competent cells. Plasmids were then sequenced to ensure the oligos were cloned into the backbone correctly.

Production of lentivirus was performed in 293T cells by co-transfection of pVSVg (Addgene #8454) and psPAX2 (Addgene, #12260) and the lentiviral plasmid using Lipofectamine 2000 (Invitrogen), as described (Sanjana et al., 2014, Shalem et al., 2014). The cell culture medium was collected 48 hours after transfection, filtered through a 0.22 μM membrane and centrifuged at 1500 g for 10 mins. The supernatant was aliquoted and stored at -80 °C.

The MCF-7 cell line was infected with the lentivirus. Puromycin (1 μg/ml) was added to the culture medium 48 hours after the infection. The cells were plated at a low density to get well separated colonies. Single cell clones were manually picked and expanded for validation.

Table 4: Primers used in this dissertation

Primer	Sequence (5' to 3')
PLAA Exon1-targeting sgRNA (F)	CACCGGGCCACGAGCTGGACGTACG
PLAA Exon1-targeting sgRNA (R)	AAACCGTACGTCCAGCTCGTGGCCC
PLAA KO Sequencing Primer (F)	GCTGAGGCCGATGATGAAT
PLAA KO Sequencing Primer (R)	GACTCGTCTGTGGTCAAGTTAG
PLAA WD40 N-term (F)	TAAGTCGACATGACGAGCGGCGCAACC
PLAA PFU N-term (F)	TTAGCTCGAGCTCCTGGTACTAGAGAAGGA
PLAA PFU C-term (R)	ATAGGATCCACTACTACCACTAGAACCCGG
PLAA PUL C-term (R)	GCCCGCGGCTACAGCAAATTTAGGATAAA
VRK1 K4A (F)	GCCTCGTGTAGCAGCAGCTCAAG
VRK1 K4A (R)	ATGGTGGCCTCGAGATCT
VRK1 TEV (F)	TTTTCAGGGCAGCTCTGCAAAGAGACATC
VRK1 TEV (R)	TACAGGTTTTCCTGTCTTCCAGCTTGAGC
BCL11B K4A (F)	GTCCCGCCGCGCACAGGGCAAC
BCL11B K4A (R)	GTCCCGCCGCGCACAGGGCAAC
TSHZ1 K4A (F)	GCCGAGGAGGGCGCAGCAGGCC
TSHZ1 K4A (R)	GCCGAGGAGGGCGCAGCAGGCC
MEF2C K4A (F)	GGGGAGAAAAGCGATTCAGATTACGAGGATTATG
MEF2C K4A (R)	GGGGAGAAAAGCGATTCAGATTACGAGGATTATG
EGFP-N1_ Sequencing Primer (F)	CGCAAATGGGCGGTAGGCGTG

Plasmid construction: Human codon-optimized plasmids, GFP-PLAA, GFP-ANKIB1, VRK1-GFP, BCL11B-GFP, MEF2C-GFP, TSHZ1-GFP, PCLAF-GFP, MAPK8-GFP were synthesized by Biomatik (Ontario, Canada). Construction of GFP-tagged PLAA truncations were done by polymerase chain reaction (PCR) using GFP-PLAA as a template. PCR uses a combination of specific primers that are added to the template DNA to amplify genes in a specific way (See **Table 4** for primer sequences).

***In vitro* methylation assay:** The GST-Set7/9 was expressed and purified as described previously (Cheng et al., 2007). The recombinant MLL1 and PRDM9 proteins were purchased from Reaction Biology Corp (Malvern, PA) and Active Motif (Carlsbad, CA), respectively. *In vitro* methylation reactions were performed in a final volume of 30 μl of 50 mM Tris-HCl (pH 8.5), 5 mM MgCl2, 4mM DTT, and 0.42 μM ^{3}H-labeled S-adenosyl-L- [methyl ^{3}H] methionine (PerkinElmer Life Sciences) (Nishioka et al., 2002, Hamidi et al., 2018).

Live cell microirradiation imaging and immunofluorescence microscopy

Laser microirradiation imaging: Laser microirradiation and quantification of live-cell imaging of laser damage recruitment were performed with Zeiss Zen software on a Zeiss LSM 880 confocal microscope (Zeiss) following standard protocols designed in previous studies (Gong et al., 2015). In brief, cells were seeded onto (World Precision Instruments, FD35-100) and co-transfected using PEI with 2 μg GFP-PLAA plasmid and 2μg RFP-53BP1. Media was changed 4~6 hours after transfected cells and cells were cultured for another 24 hours. All cells used for laser microirradiation and after the experiments (live-cell imaging and immunofluorescence) were pre-sensitized by adding 10 μM BrdU in regular DMEM medium for 20 hr or with 200ng/mL Hoechst33342 (Thermo Scientific, Cat# 62249) for 30 mins before damage treatment. Live-cell imaging experiments were performed at 37°C and 5% CO2

conditions maintained by a heated incubation system on the microscope. All images were captured using a Zeiss 40X water objective lens. A 355-nm laser beam (30%) was used to generate laser microirradiation. For each experiment, quantification data were collected from >10 cells under each condition. Plots shown in figures are from one representative experiment.

Immunofluorescence: After indicated treatments, cells were rinsed with 1X PBS and then fixed/permeabilized with 100% Methanol at -20 °C for 10 mins (as required by H2Bub1 antibody) or fixed with 4% formaldehyde for 10 min at room temperature for all other antibodies. Upon fixation, cells were blocked with PBS containing 3% BSA. After blocking, cells were incubated with indicated primary antibodies and corresponding secondary antibodies. Finally, the cells were stained with DAPI at room temperature. Confocal images were acquired using a Zeiss LSM 880.

CHAPTER 3: PROTEIN DOMAIN MICROARRAY AS A PLATFORM TO DECIPHER SIGNALING PATHWAYS AND THE HISTONE CODE

Part of this chapter is based upon: J. Chen, C. Sagum, and M.T. Bedford, (2020) Protein domain microarrays as a platform to decipher signaling pathways and the histone code. Methods, Volume 184: pp. 4-12.

ISSN 1046-2023

3.1 CELL SIGNALING THROUGH POSTTRANSLATIONAL MODIFICATIONS

PTM is the process of covalently adding a chemical group to an amino acid in a protein after the protein has been synthesized. PTMs are usually enzymatically added to and removed from the protein substrates. The most prominent types of PTMs are phosphorylation, methylation, acetylation, glycosylation and ubiquitination, but this is by no means a complete list (Deribe et al., 2010a). It has long been known that PTM patterns are altered in response to extracellular signals like growth stimulation and intracellular signals like DNA damage. Thus, activated and altered signaling cascades drive information from the cell surface to the nucleus, and also from the nucleus to the cytoplasm. These transduction processes are critical for cells within organisms to respond to information in their environment, and often culminate in epigenetic changes and altered gene expression signatures (Badeaux and Shi, 2013). These signals are transduced or relayed by effector proteins that often read one type of PTM and deposit a different type of PTM, thus the term “signal transduction cascade”.

3.2 PROTEIN DOMAINS: INTERACTION HUBS THAT DRIVE SIGNALING CASCADES

The concept that distinct protein regions within a continuous peptide chain can form globular structures was originally proposed by Donald Wetlaufer (Wetlaufer, 1973). This notion of globular protein domains was expanded to include both structure and function (Richardson, 1981). Domains are the smallest functional and structural unit of a protein and are usually between 30 to 100 amino acids in length. Protein domains are not only regions that are expected to be stable as independent units, but they are also often conserved regions between different proteins. Thus, different proteins can be grouped into a family if they share the same domain, as the shared domain often confers similar functions. Protein domain identification has been greatly improved with advances in bioinformatics techniques. This provides the premise to predict novel domains/functions of a protein based on both sequence and structural information. The discovery by the Pawson group in the mid 90s, that SH2

domains are the modular units that engage phospho-tyrosine motifs (Songyang et al., 1993), set the stage for the discovery of a large number of domain types that “read” different PTMs, unique motif folds (like proline-rich sequences), phospholipids, metabolites and modified nucleic acids (**Figure 1 & Table 5**). Therefore, domains can be considered the structural “hubs” within proteins that drive interactions, and facilitate the formation of regulatable protein interaction networks and propel signal transduction.

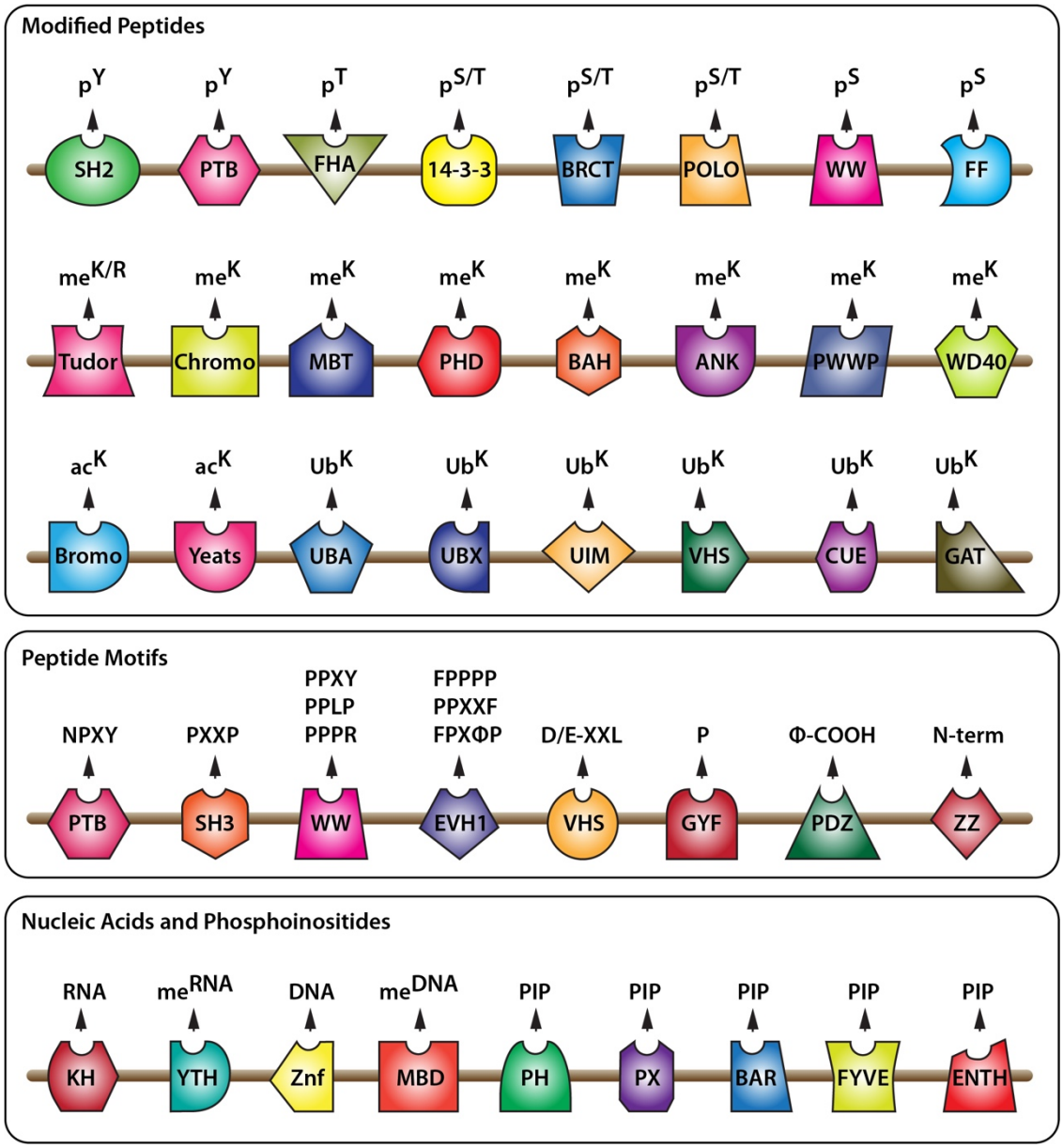


Figure 1. Distinct domains, depicted with different colors and shapes are able to recognize specific motifs. Some domains recognize motifs that harbor PTMs, while others recognize distinct proline-rich sequences, nucleic acids or phospho-lipids. Y, tyrosine; S, serine, T, threonine; K, lysine; P, proline; ϕ-COOH, hydrophobic C-terminal; p, phosphorylation; me, methylation; ac, acetylation; ub, ubiquitination; PIP, phosphoinositides.

Table 5. List of major protein domains that participate in signal transduction

SH2	Src Homology 2
PTB	Phosphotyrosine-binding domain
FHA	Forkhead-associated domain
14-3-3	The 14th elution fraction of DEAE-cellulose chromatography and the migration position 3.3 of subsequent electrophoresis
BRCT	BRCA1 C Terminus domain
PBD	Polo-box domain
WW	Containing two conserved Tryptophans that are 20-22 amino acids apart
Tudor	Named after the Tudors family for grandchildless phenotype
Chromo	chromatin organization modifier domain
MBT	Malignant brain tumor domain
PHD	Plant homeodomain
BAH	Bromo-adjacent homology
ANK	Ankyrin repeats domain
PWWP	Presence of a central core region 'Pro-Trp-Trp-Pro'
WD40	A 40 amino acid motif with a terminating ‘Trp-Asp’ dipeptide
Bromo	Identified as a novel structural motif when studying the drosophila gene Brahma (brm)
YEATS	Yaf9, ENL, AF9, Taf14, Sas5 domain
UBA	Ubiquitin-associated domain
UBX	An 80 amino acid module with unknown function found in ubiquitin-related/unrelated proteins
UIM	Ubiquitin-interacting motif domain
VHS	Originally found in VPS-27, Hrs and STAM proteins
CUE	A homology region between yeast Cue1 and human Tollips proteins
PFU	PLAA family ubiquitin-binding domain
GAT	A homology region in eukaryotic GGAs and vertebrate TOMs proteins
ZZ	ZZ-type zinc finger (ZZ)
PTB	Phosphotyrosine-binding domains
SH3	Src Homology 3
EVH1	WH1, RanBP1-WASP, or enabled/VASP homology 1 domain

3.3 EARLY DEVELOPMENT OF PROTEIN MICROARRAYS

Some of the earliest protein screening approaches employed phage expression libraries to identify novel protein-protein interactions (Cicchetti et al., 1992, Chan et al., 1996) and discover kinase substrates (Fukunaga and Hunter, 1997). High-density arrays using bacterially produced His-tagged proteins were also developed at about the same time, and were used for protein ligand screening (Bussow et al., 1998) and enzyme substrate screening (Lee and Bedford, 2002). The disadvantage of both approaches was that the recombinant proteins were arrayed on large membranes, and not microarrayed on slides, which made certain assays difficult, because enzymatic reactions or protein interaction probing experiments need to be performed in very large volumes (30-50 ml). This issue spawned the development of microarrays that could be used for miniaturized high-throughput screening, in which proteins were spotted onto chemically derivatized glass slides at high density (MacBeath and Schreiber, 2000, Zhu et al., 2001). More recently, large-scale protein domain microarrays have become broadly available, and represent the commercialization of pioneering work from Snydner's group at Stanford University (Smith et al., 2005) and the Zhu and Blackshaw laboratories at Johns Hopkins University (Jeong et al., 2012, Lu et al., 2013). The ProtoArray™ is offered by Thermo Fisher Scientific and contains over 9,000 proteins on a single slide. These human proteins are expressed in insect cells as N-terminal GST-tagged fusions and printed onto nitrocellulose coated glass slides. The second commercially available human proteome microarray is called HuProt™ and is available from CDI. This microarray harbors over 16,000 proteins, which are also N-terminal GST-tagged fusions, but in this case, the recombinant proteins are expressed in yeast cells and then printed directly onto glass slides. A third type of protein array was generated by spotting protein-coding plasmids onto slides, and then translating these coding regions into proteins using a cell free reticulocyte lysate (He and Taussig, 2001, Ramachandran et al., 2004). These proteins are fused to a tag that allows immediate immobilization *in situ*, and they are called nucleic acid

programmable protein arrays (NAPPAs). Human NAPPA arrays with over 11,000 proteins are available through the Protein Array Core of the BioDesign Institute in Arizona.

These three large human proteome array types - ProtoArray™, HuProt™ and NAPPAs – provide a valuable resource for the screening of enzyme substrates and high-affinity protein-protein interactions, and for epitope mapping. However, these array types have proven more difficult to use for the screening of weaker protein-protein interactions that are driven by PTMs, which generally display binding constants in the low μM range. This issue can be mitigated to some degree by arraying much higher concentrations of the protein region that interacts with the PTM motif. Furthermore, when making the recombinant proteins for the content of microarrays, there is the concern that the proteins will not express well in bacteria, yeast or insect cells, especially when attempts are made to generate full-length proteins that can be rather large and prone to precipitation and inclusion body formation. This concern can largely be alleviated by using protein domains. The nature of protein domains (small, well-expressed, tightly folded and functionally stable) make them ideal for expression as recombinant proteins that can be used for multiplexing. We were among the first groups to develop and use protein domain microarrays in the early 2000s (Espejo et al., 2002, Espejo and Bedford, 2004b), and their use has rapidly evolved over the past two decades. At MD Anderson Cancer Center we have developed a core facility – the Protein Array and Analysis Core (PAAC) – to generate libraries of different recombinant domains and array these libraries for discovery purposes.

3.4 THE DESIGN OF PROTEIN DOMAIN MICROARRAY FOR THE PAAC

The first step in generating protein domain microarrays is identifying all the domain types that need to be generated for a particular project. This is done with the aid of web-based programs that can predict the presence of protein domains in the proteins of interest, including EMBL *Pfam* and *InterPro*, NCBI *Conserved Domain* and the CST *PhosphoSite*. Often, multiple domains of the same type or of different types will lie close together. If this is the case, we

select these different domains together, if they are within 50 amino acids of each other. The second step is to design and generate expression vectors for all of the domains that will be included on the arrays. Usually, we add 20 amino acids of flanking sequence to each domain, and then order commercial gene synthesis and cloning of the domain. We use a Canadian company, Biomatik, for this step. Importantly, all the domains that the PAAC designs are from human proteins, and for the gene synthesis, we codon optimize these human sequences for bacterial expression. Once the expression vectors are in-hand, the third step is to generate the recombinant protein domains that will serve as the content for the array. Protein domains of interest are made as N-terminal GST fusions. We do not use a high-throughput approach for recombinant GST-domain production. Finally, the purified recombinant protein domains are robotically arrayed onto nitrocellulose-coated glass slides (Oncyte®Avid slides, Grace Bio-Labs, Bend, OR), using an Aushon 2470 pin microarrayer.

The recombinant proteins in the “source” plate are at a concentration of 1µg/µl, and we have empirically determined that repeated arraying (five times) onto the same spot results in a strong signal, while retaining the integrity of the spotted protein. We are not sure how much protein is actually immobilized on the array. Duplicates of each protein domain are used to facilitate identification and reproducibility. Once an array is generated it is stored at 4 °C, and a probe is then prepared. The arrays can be stored for at least one month at 4 °C without losing their binding specificity. In our hands, long-term storage of arrays for up to a year at -80 °C has also been successful.

3.5 PROBE PREPARATION FOR PROTEIN DOMAIN MICROARRAY INTERROGATION

Protein domain microarrays can be probed with peptides, full-length tagged proteins, protein complexes, recombinant nucleosomes, small molecules, DNA and RNA. The probes are

usually biotinylated, which facilitates labeling, but epitope tagged proteins can also be used. The protein domain microarrays are primarily used to identify protein interactions that are driven by PTMs, and for these cases we use biotinylated peptide probes. Peptide probes are designed to have eight residues on either side of the amino acid that is carrying the PTM (thus 17-mer peptide) with a biotin moiety attached to either the C- or N-terminal residue. If a small molecule is used as a probe, the compound of interest is synthesized with a PEG linker and biotin. A streptavidin-fluorophore (usually Cy3 or Cy5) conjugate is then coupled to the biotinylated probe. The array slides are then blocked, probed and washed. This probing is performed in a small slide holder in a volume of 1.5 ml. In this volume we use 50-100µg of peptide (or recombinant protein) as a probe. The probed slides are then air dried and the fluorescent signal is detected using a GenePix 4200A Microarray Scanner (Molecular Devices). The workflow is shown in **Figure 2,** and additional details have been published (Espejo et al., 2002, Espejo and Bedford, 2004b).

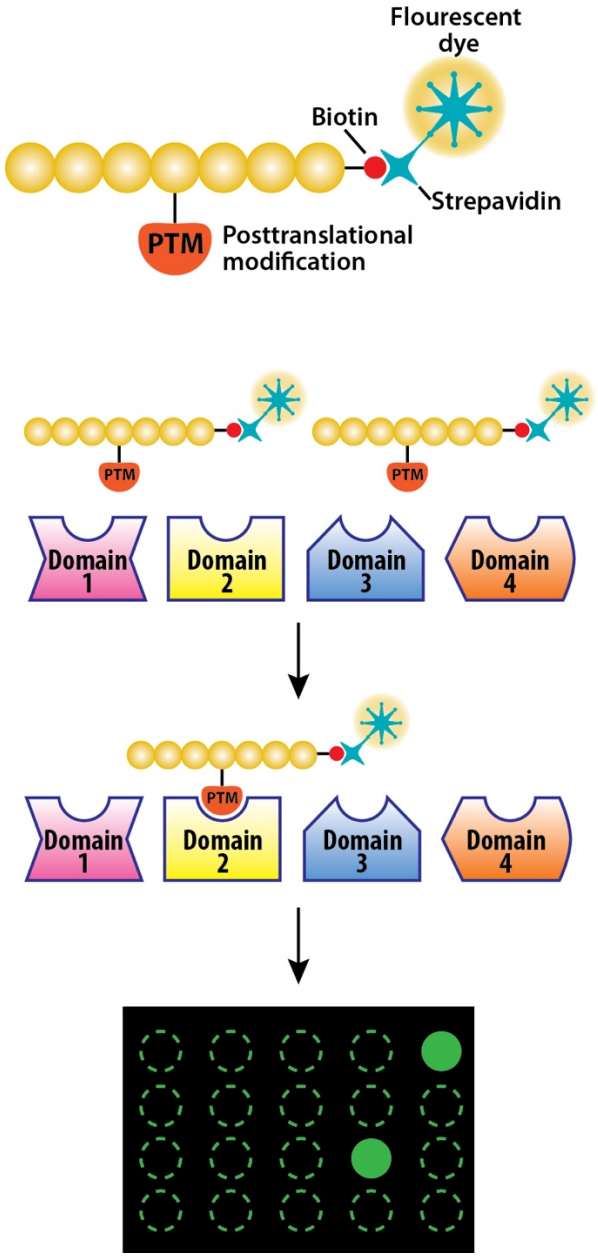


Figure 2. Principles for protein domain microarray screening. Different domains are immobilized on a nitrocellulose-coated glass slide as GST fusion proteins. Biotinylated peptides (100μg) with specific PTMs are pre-conjugated with streptavidin fluorophores. Peptide probes are incubated with the microarray followed by washes to remove unbound materials. Interactions are visualized by detecting fluorescent signals. Within a block, we array a duplicate of each protein, and each pair of spotted proteins is arrayed at a unique angle (not adjacent to one another) to facilitate rapid unambiguous identification.

3.6 USING DOMAIN MICROARRAY TO IDENTIFY NOVEL PROTEIN INTERACTIONS

3.6.1 Domain arrays that "read" Ubiquitin

Ubiquitin is a highly conserved 76 amino acid protein that can be attached (though its C-terminal glycine residue) to lysine residues in substrates by an enzyme cascade. The reversible attachment of a single ubiquitin (monoubiquitylation) was initially discovered forty years ago as a modification of histone proteins (Goldknopf et al., 1977), where it controls chromatin structure and regulates transcription rates. Ubiquitin itself harbors seven lysine residues that can all serve as attachment sites for further ubiquitin molecules, thereby forming polyubiquitin chains (Yau and Rape, 2016). There are at least twenty distinct families of structurally unique ubiquitin binding domains (UBDs) (Dikic et al., 2009). UBDs bind ubiquitin with relatively low affinity (10-500 μM range). However, UBDs are often repeated in proteins, which can result in an avidity effect that promotes strong interactions. Furthermore, long ubiquitin chains provide the perfect scaffold for the binding of multiple UBD-containing proteins and the rapid assembly of ubiquitin chain driven signaling nodes. Moreover, the varying topologies of the different chains provide distinct binding surfaces for UBDs. The first UBD to be identified and characterized was a domain in the S5a protein, which is a component of the 26S proteasome (Deveraux et al., 1994). S5a has two independent ubiquitin-interacting motifs (UIMs), and bioinformatics analysis uncovered this domain in a number of other proteins. UIMs thus form the founding family of UBDs, but many more structurally distinct UBD families have now been identified. The second largest family of UBDs to be identified was the ubiquitin-associated domains (UBAs) (Bertolaet et al., 2001, Wilkinson et al., 2001). Subsequent screens using the yeast two-hybrid approach with ubiquitin and ubiquitinated proteins as bait, identified the CUE domain (Shih et al., 2003), GAT domain (Shiba et al., 2004), VHS domains (Yamakami et al., 2003) and PAZ domains (Hook et al., 2002). Other approaches identified PFU, UBM, UBZ and JAB/MPN domains (Fu et al., 2009b, Bienko et al., 2005, Bellare et al., 2008) as effectors of ubiquitinated proteins. Structural

studies have revealed that different surfaces of ubiquitin are targeted for non-covalent interactions by different UBDs, with the most common interaction surfaces centered on a hydrophobic patch around Ile44, an acidic region around Asp58, and Gly76 at the C-terminus (Scott et al., 2015). Importantly, some UBDs recognize sequences adjacent to the isopeptide linkages, which differ for each chain type, and thus generate chain selective UBD binders. Another way to achieve ubiquitin-linkage specific binding is for a protein to harbor tandem UBDs that are spaced for optimal engagement of one chain type and not another (Sims and Cohen, 2009). Additionally, PTMs on ubiquitin itself can regulate interactions with UBDs (Yau and Rape, 2016). The PAAC has recently generated an array of 169 recombinant ubiquitin-binding domains, which is being used to identify 1) the readers of mono-Ub marks, 2) chain-specific readers, and 3) the role that phosphorylation and acetylation of ubiquitin itself plays on regulating UBD interactions.

3.6.2 Domain arrays that "read" methylated motifs

Whereas the 1990s was the decade for identifying domains that "read" phospho-marks, the first decade of the new millennium was dedicated to the discovery of readers of the histone code. The histone code hypothesis was proposed by Strahl and Allis in 2000 (Strahl and Allis, 2000b) and refined a year later to include the concept of effector proteins or readers (Jenuwein and Allis, 2001b). The Jenuwein lab identified Chromo domains as the first readers of methyllysine marks on histone tails (Lachner et al., 2001). This led to a flurry of activity in the field, with the subsequent descriptions of Tudor domains (Huyen et al., 2004, Huang et al., 2006b), WD40 repeats (Wysocka et al., 2005), PHD domains (Shi et al., 2006), MBT domains (Kim et al., 2006b, Trojer et al., 2007), PWWP domains (Wang et al., 2009b), BAH domains (Kuo et al., 2012) and ankyrin repeats (Collins et al., 2008) as methyl-mark readers. During this period, we developed a protein domain microarray to rapidly read the histone code (Kim et al., 2006b). We have also used this array approach to discover TDRD3 as a Tudor

domain effector of methylarginine marks on histone tails (Yang et al., 2010). Similar array approaches were developed to identify methyl-mark readers in yeast (Shi et al., 2007) and plants (Zhao et al., 2017). These methyl-reading arrays have also been used to identify readers of non-histone proteins (Levy et al., 2011, Zhang et al., 2016). The PAAC has generated a methyllysine binding array that contains 308 potential Kme reader domains, including 33 chromodomains, 43 Tudor domains, >100 PHD, as well as representatives from a number of additional domains like PWWP, BAH, ELM2, HORMA domains, and Ank repeats.

CHAPTER 4: INVESTIGATION OF PLAA AS A NOVEL H2BUB1 READER PROTEIN

4.1 INTRODUCTION

Histone H2Aub1 and H2Bub1 are two critical epigenetic marks with opposite transcription regulation functions and they play distinct roles in DNA damage response pathways. There are two known H2Aub1-specific reader proteins, RSF1 and JARID1, but no H2Bub1 effectors have yet been identified. Thus, the effector modules' identities for these two marks remain largely unknown. The identification and characterization of specific effectors are essential to understand the mechanisms of how H2Aub1 and H2Bub1 transduce the molecular signals to their downstream protein complex. In the current study, I hypothesize that there are additional specific reader proteins for H2Aub1 and H2Bub1 that have yet to be discovered. Using an in-house developed ubiquitin-binding domain microarray, I started the project by searching for specific reader proteins for either H2Aub1 or H2Bub1 and characterizing their potential functions as effectors of these two epigenetic marks.

4.2 RESULTS AND DISCUSSIONS

Identification of novel H2Aub1- and H2Bub1-specific reader proteins

To identify potential H2Aub1- and H2Bub1-specific reader proteins, we used the H2Aub1 and H2Bub1 probes to interrogate an in-house ubiquitin-binding domain microarray that harbors GST fusions of known and predicted ubiquitin binders. By comparing the binding profiles, HRS/HGS (UIM domain) was identified as a potential H2Aub1-specific domain, and PLAA (PFU domain) and ANKIB1 (UIM domain) were two candidates for H2Bub1-specific binding (**Figure 3**).

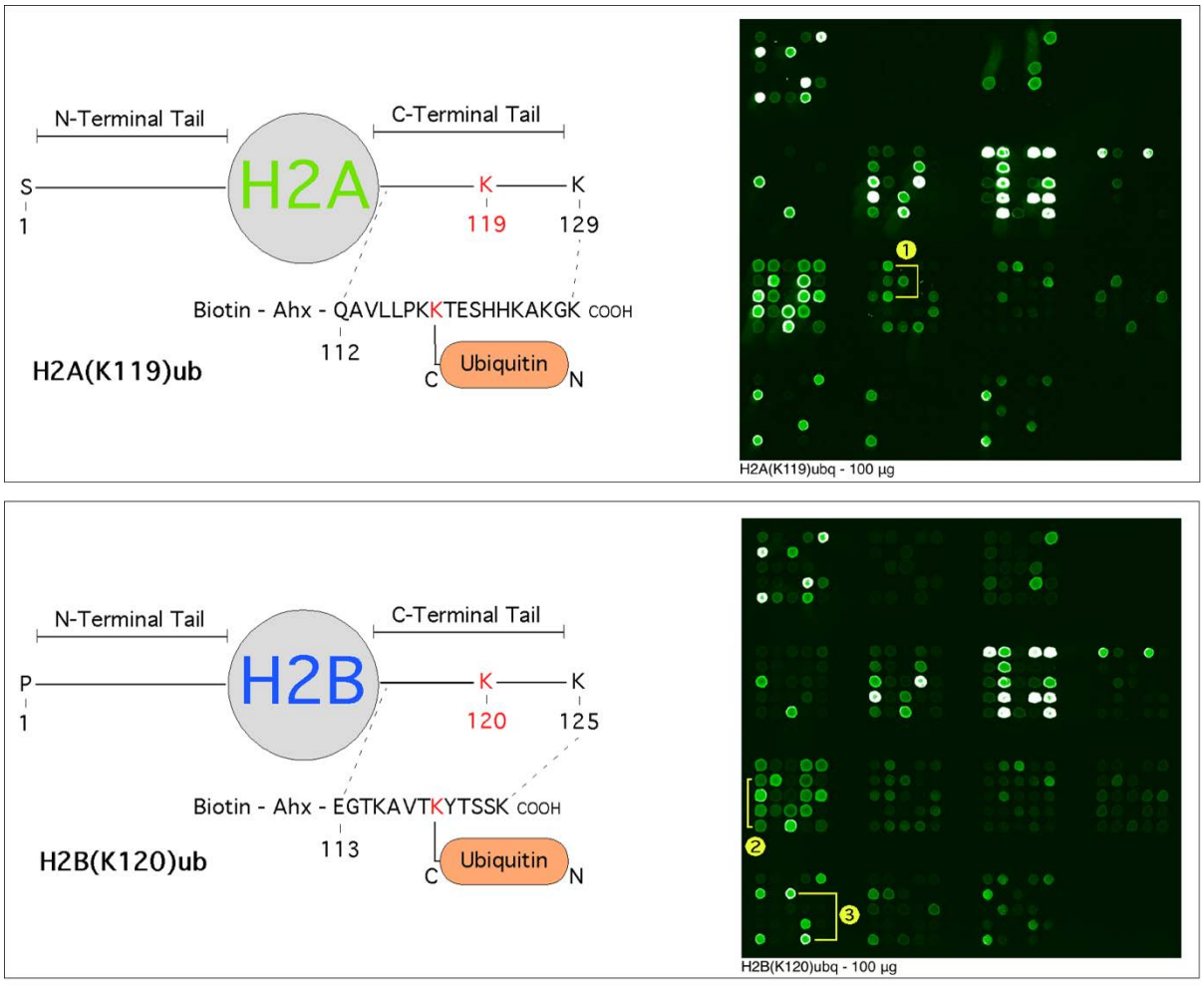


Figure 3: Identification of novel H2Aub1- and H2Bub1-specific reader proteins using ubiquitin binding domain microarray. Top: H2Bub1 probe structural information (left) and the UBD binding profile (right). HGS/HRS UIM domain was highlighted with a circle number 1. Bottom: H2Bub1 probe structural information (left) and binding profile with UBDs (right). ANKIB1 UIM and PLAA PFU domains were highlighted with circled number 2 and 3, respectively.

Validation of novel interactions using *in vitro* binding assays

Cari Sagum performed a peptide pull-down assay to confirm the interactions identified using the protein domain microarrays. H2Aub1 and H2Bub1 were used to pull down the GST fusions of ubiquitin-binding domains from HRS/HGS, ANKIB1, PLAA, UBAC2 and AMFR. GST served as a negative control. UBAC2 and AMFR served as positive controls as both domains interacted with H2Aub1 and H2Bub1 probes in our microarray results. Consistent with the array data, HRS/HGS UIM interacted with H2Aub1 but not H2Bub1, while ANKIB1 UIM domain and PLAA PFU domain specifically interacted with H2Bub1 but not H2Aub1 probe (**Figure 4A**).

The next step was to test whether GST-fused UBDs could preferentially interact with full-length H2Aub1 or H2Bub1 from acid purified histone extracts. I first validated H2Aub1 and H2Bub1 antibodies' specificity using the ATXN7L3 knockdown (KD) cell lines, generously provided by Dr. Sharon Dent's group. ATXN7L3 is the adapter protein that activates three H2Bub1 deubiquitinating modules, USP22, USP51, and USP27X (Atanassov et al., 2016). Depletion of ATXN7L3 leads to increased global H2Bub1 levels without changing H2Aub1 levels. Consistent with published results (Atanassov et al., 2016), the H2Bub1 antibody detected significantly increased signals in ATXN7L3 knockdown cells but no changes of H2Aub1 levels compared to the control cells. **(Figure 5A**).

I then tested whether the three GST-fused UBDs could pull down their preferential ubiquitinated histones from acid purified histones. HRS/HGS UIM domain didn't enrich H2Aub1 nor H2Ub1 from histone extracts thus was excluded from the following study. ANKIB1, PLAA, AMFR, and UBAC2 all displayed comparable binding properties towards H2Aub1 and H2Bub1, in this pull-down assay. All four UBDs interacted more strongly with H2Bub1 than H2Aub1, as compared to the input signals (**Figure 5B**).

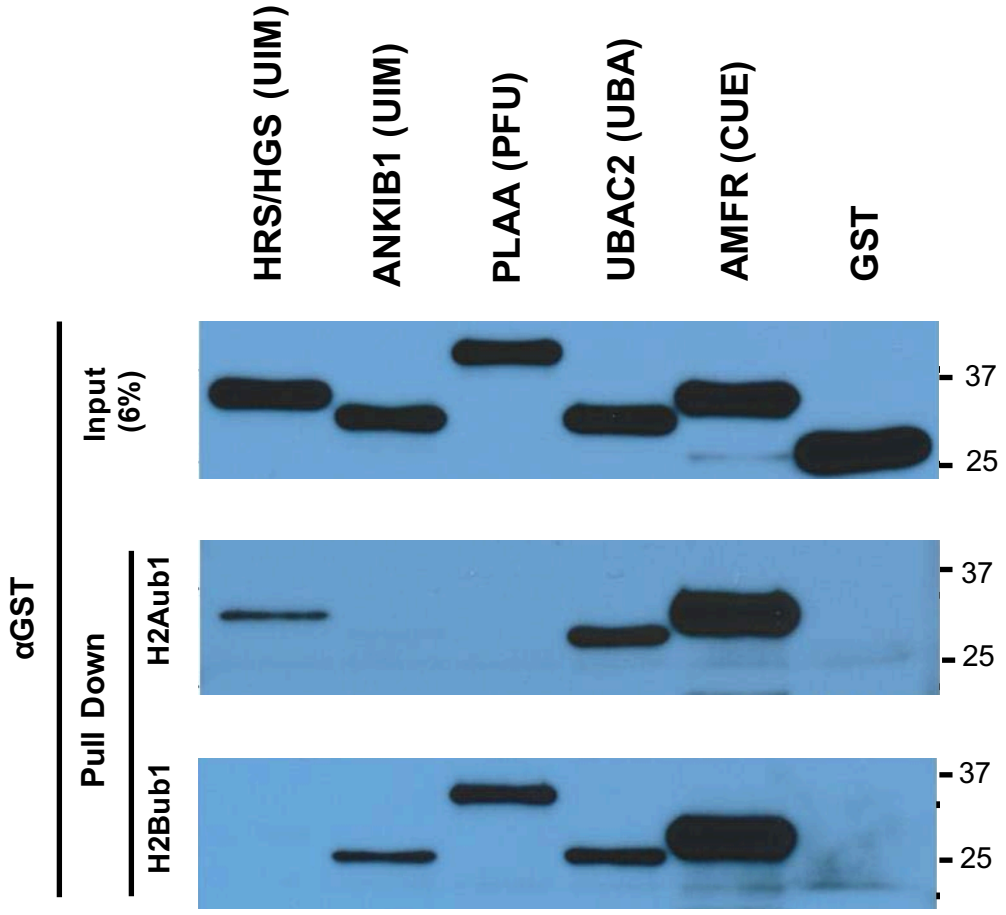


Figure 4: Validation of novel Interactions using peptide pull-down assays. This data was generated by Cari Sagum and is used with her permission. Top panel: 5% input of all tested GST proteins. Middle panel: HRS/HGS, UBAC2, and AMFR were pulled-down by the H2Aub1 peptide probe. Bottom panel: H2Bub1 probe pulled down ANKIB1, PLAA, UBAC2, and AMFR proteins.

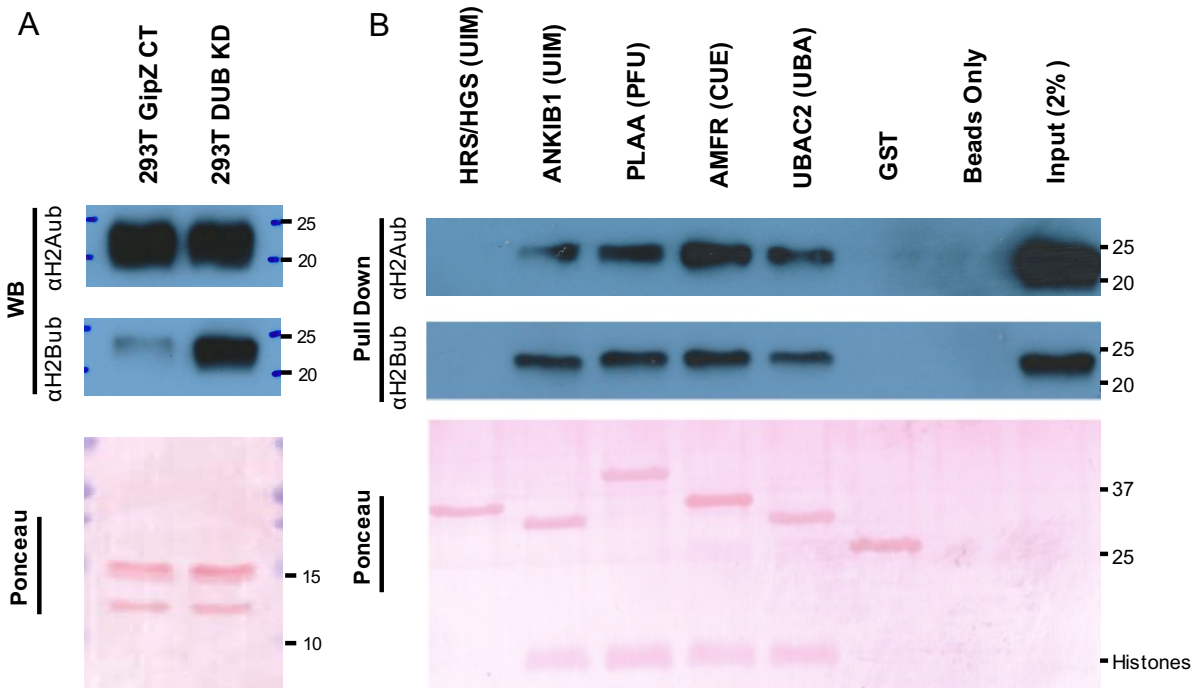


Figure 5: Characterization of H2Bub1 antibody and validation of novel interactions using pull-down assays. (A) Histones were acid purified from ATXN7L3 KD, and control (CT) cells were separated by SDS-PAGE and followed by western blot analysis for H2Aub1 and H2Bub1 levels. Ponceau staining serves as a loading control. 293T GipZ CT: control cell lines transduced with the Non-Silencing Control Vector. 293T DUB KD: cell lines transduced with ATNX7L3 shRNA lentiviral vector. (B) GST-fused UBDs were used to pull down H2Aub1/H2Bub1 from acid purified histone extracts. GST and Beads only served as negative controls while AMFR and UBAC2 served as positive controls.

Characterization of anti-PLAA and anti-ANKIB1 antibodies

To facilitate experiments in cells, I characterized the antibodies against PLAA and ANKIB1 using overexpression of GFP-fused PLAA and ANKIB1 in MCF-7 cells. As shown in **Figure 6**, both antibodies showed excellent specificity by recognizing the GFP-tagged and endogenous PLAA and ANKIB1. Ectopic expression of GFP-PLAA is significantly more abundant than endogenous PLAA, a 78 kDa protein labeled by a red arrow.

I also examined the subcellular localization of PLAA and ANKIB1 using immunofluorescence microscopy. As shown in **Figure 7**, PLAA mainly exists in the nucleus and less in the cytosol. ANKIB1 is localized in the cytosol. Considering they are potential reader proteins for H2Bub1, a nuclear modification, ANKIB1 was excluded in the subsequent studies.

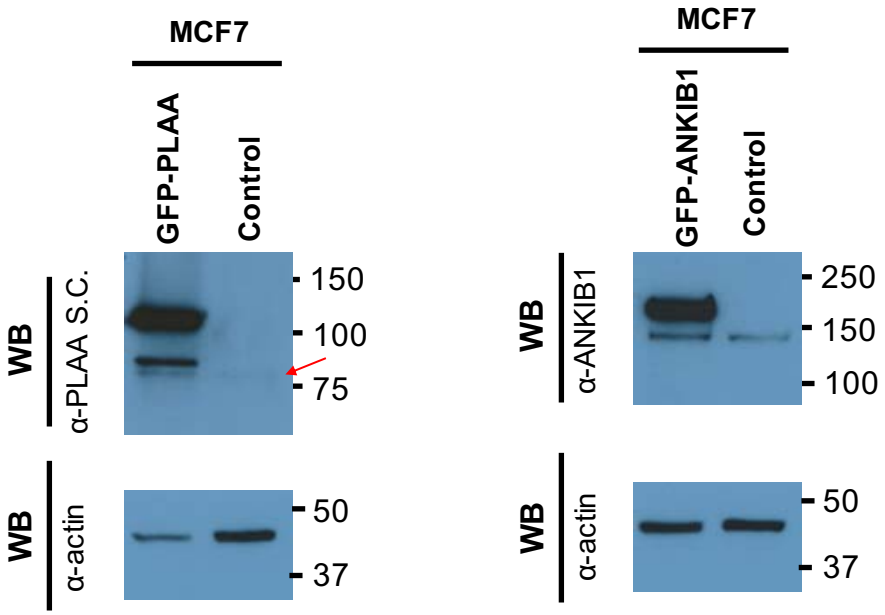


Figure 6: Characterization of PLAA and ANKIB1antibodies using ectopic expression of GFP fusion proteins. GFP-PLAA and GFP-ANKIB1 were overexpressed in MCF-7 cells. Untransfected cells serve as negative controls. Whole-cell lysates were separated by SDS-PAGE followed by western blot analysis against PLAA, ANKIB1, and beta-actin. The red arrow indicates the endogenous PLAA (75 kDa).

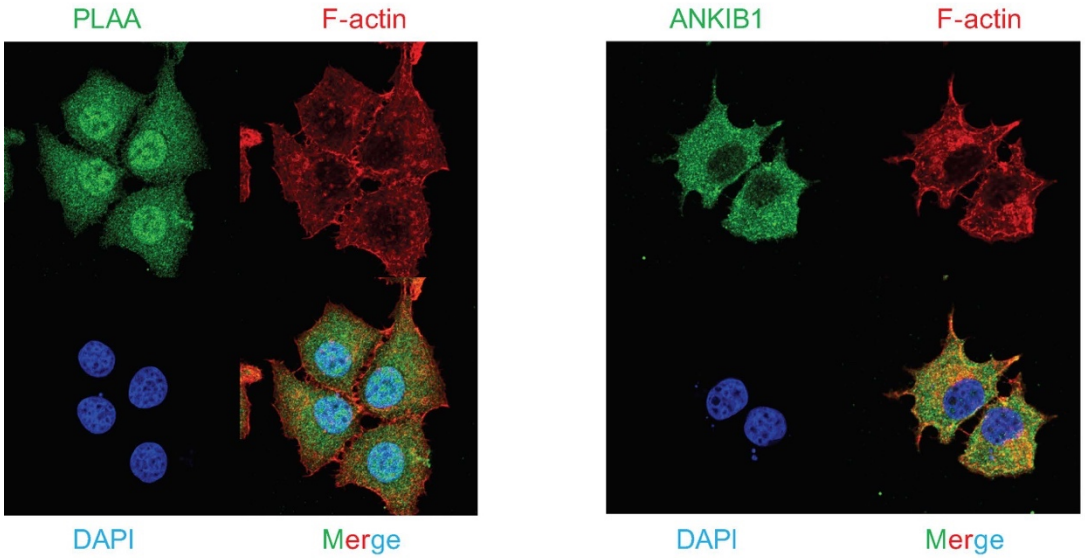


Figure 7: Sub-cellular localization of PLAA and ANKIB. MCF-7 cells were fixed with 4% formaldehyde, and stained for PLAA, ANKIB1, F-actin and DAPI. Immunofluorescence was detected using a Carl Zeiss LSM880 microscope equipped with an Aryscan confocal microscope (Zeiss, Jena, Germany)

Validation of PLAA-H2Bub1 interaction using nucleosome binding assays

Recombinant nucleosomes have become widely used to characterize epigenetic writers, readers, and erasers, in biochemical and structural studies (Kim et al., 2020, Chen et al., 2015, Zhang et al., 2017a). Testing the binding potential of nucleosomes is important because in some cases effector proteins can read short linear motifs and even free histones, but are unable to dock onto histones in the context of tightly packed nucleosomes, due to structural interference from other histone or DNA. Dr. Zu-Wen Sun from EpiCypher generously provided three types of recombinant nucleosomes, unmodified, H2Aub1, and H2Bub1 containing nucleosomes. The DNA that is wrapped around these recombinant nucleosomes is biotinylated, which serves at a tag for immobilization and detection. Before performing nucleosome binding assays, I characterized these nucleosomes using Coomassie staining and western blot. As shown in **Figure 8A**, the H2Aub1 nucleosome contains H2Aub1, H2B, H3 and H4 but no H2A, while the H2Bub1 nucleosome contains H2A, H2Bub1, H3 and H4 but not H2B, indicating that the respective nucleosomes contain fully ubiquitinated H2A and H2B. H2Aub1 and H2Bub1 were also detected by the corresponding antibodies (**Figure 8B**), confirming that three recombinant nucleosomes consist of high-quality core histones with the proper ubiquitination marks because H2Aub1 and H2Bub1 antibodies recognize the branching region where ubiquitin forms an isopeptide with the H2A at K119 and H2B at K120 (Minsky et al., 2008, Vassilev et al., 1995).

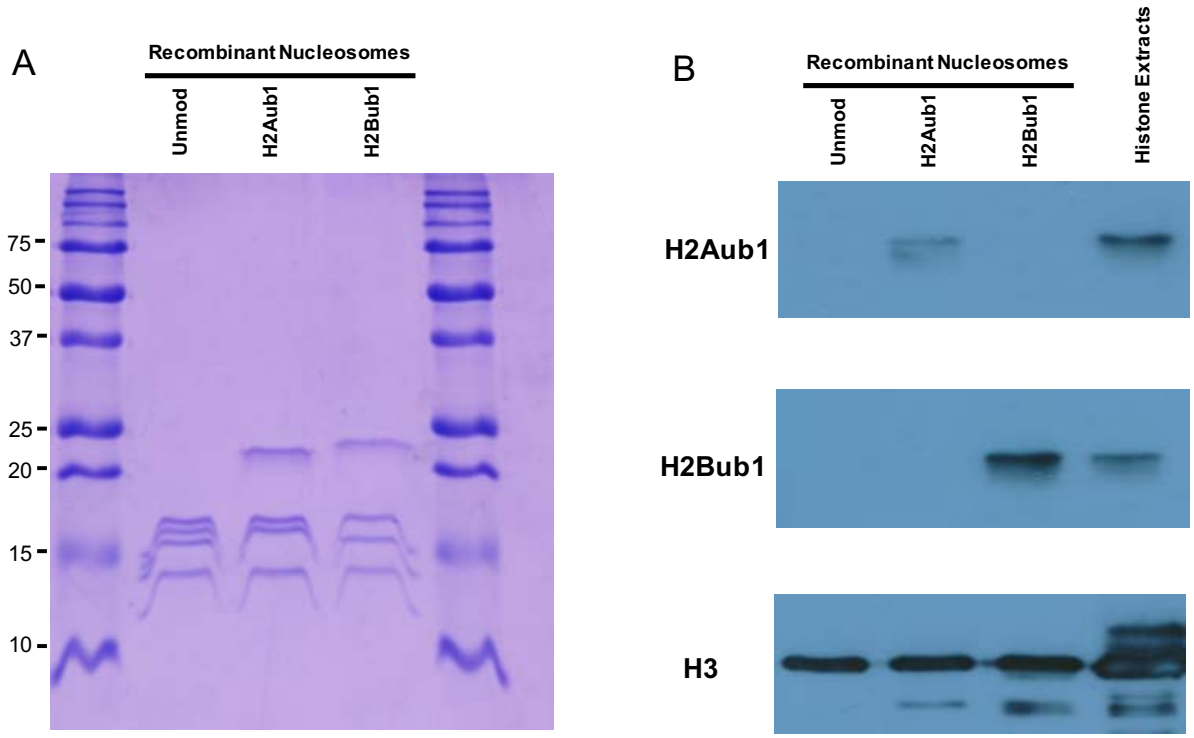


Figure 8: Characterization of recombinant nucleosomes. Three recombinant nucleosomes (unmodified, H2Aub1 and H2Ub1) were separated by SDS-PAGE and subject to Coomassie staining (A) and western blot for H2Aub1, H2Bub1 and H3 (B). Histone extracts were used as positive controls.

I then tested whether the PLAA PFU domain can specifically interact with H2Bub1 in the context of nucleosomes (**Figure 9**). GST-PFU domain pulled down H2Bub1 specifically from histone extracts but didn't interact with H2Bub1 nucleosomes, suggesting this interaction might require additional chromatin context missing in recombinant nucleosomes, or alternatively, the PFU domain interacts with free H2Bub1 but not with nucleosomal H2Bub1.

Notably, GST-PFU also pulled down unmodified H2B from histone extracts. As seen in **Figure 5B**, ponceau staining showed that unmodified core histones were pulled down together with H2Bub1 by the four UBDs.

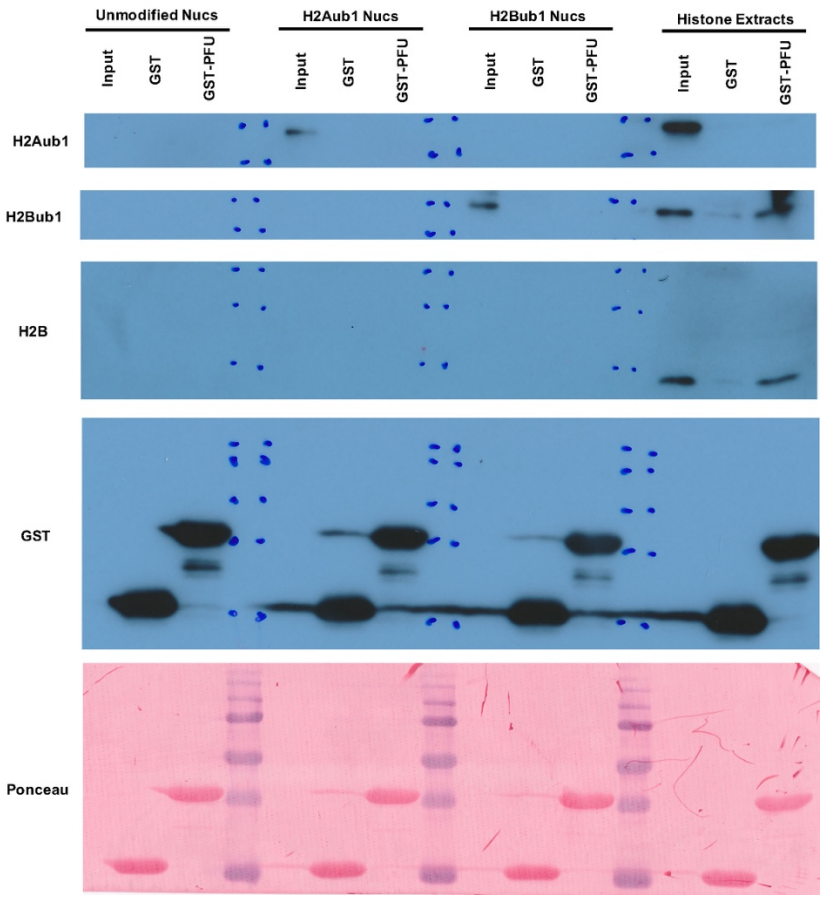


Figure 9. Testing the PLAA-H2Bub1 interaction using nucleosome binding assays. GST and GST-PFU proteins were incubated with three recombinant nucleosomes and histone extracts. GST served as a negative control, and histone extracts served as positive controls.

Generation of PLAA Knockout MCF-7 Clones

To interrogate functions of PLAA *in vivo*, I generated four single cell-derived PLAA knockout (KO) MCF-7 clones using CRISPR/Cas9. Unlike Doa1, PLAA depletion didn't affect the amount of free ubiquitin in the whole-cell extracts. Surprisingly, I observed a 15 kDa protein accumulated that can be recognized by the anti-ubiquitin antibody in PLAA KO cells and named it 15UBL protein (**Figure 10A**). Cellular fractionation assays showed 15UBL exists in the cytoplasm (**Figure 10B**). I also tested how non-specific inhibition of DUBs affected 15UBL using PR-619, which is a broad spectrum DUB inhibitor (Seiberlich et al., 2012). PR-619 treatment caused slight upregulation of 15UBL in PLAA KO cells but had no effects on WT cells, suggesting PLAA is indispensable for 15UBL's turnover by DUBs (**Figure 10C**). It remains unknown which DUBs mediate hydrolysis of 15UBL.

Western blot analyses ruled out K48- and K63-linked di-Ub, RPS27A and interferon-stimulated gene 15 (ISG15) (data not shown). RPS27A is one of two ribosomal proteins containing an N-terminal ubiquitin and a c-terminal ribosomal protein S27A(Redman and Rechsteiner, 1989). ISG15 is a ubiquitin-like protein induced upon type I interferon treatment (Korant et al., 1984, Blomstrom et al., 1986). Both RPS27A and ISG15 have molecular weights of around 15 kDa and can be recognized by the anti-ubiquitin antibodies. We are still not sure of the identity of this protein.

PLAA depletion also led to a major decrease in H2Aub1 and H2Bub1. H3K4me3, which can be promoted by H2Bub1 during transcription activation (Kim et al., 2009, Racine et al., 2012), was not affected in PLAA KO cells (Figure 11A). In yeast, Doa1 depletion abolishes PCNA ubiquitination upon DNA damage (Lis and Romesberg, 2006). However, upregulation of PCNA ubiquitination after UV-induced DNA damage treatment was not affected in PLAA KO cells (Figure 11B). Thus, this function of Doa1/PLAA is not conserved.

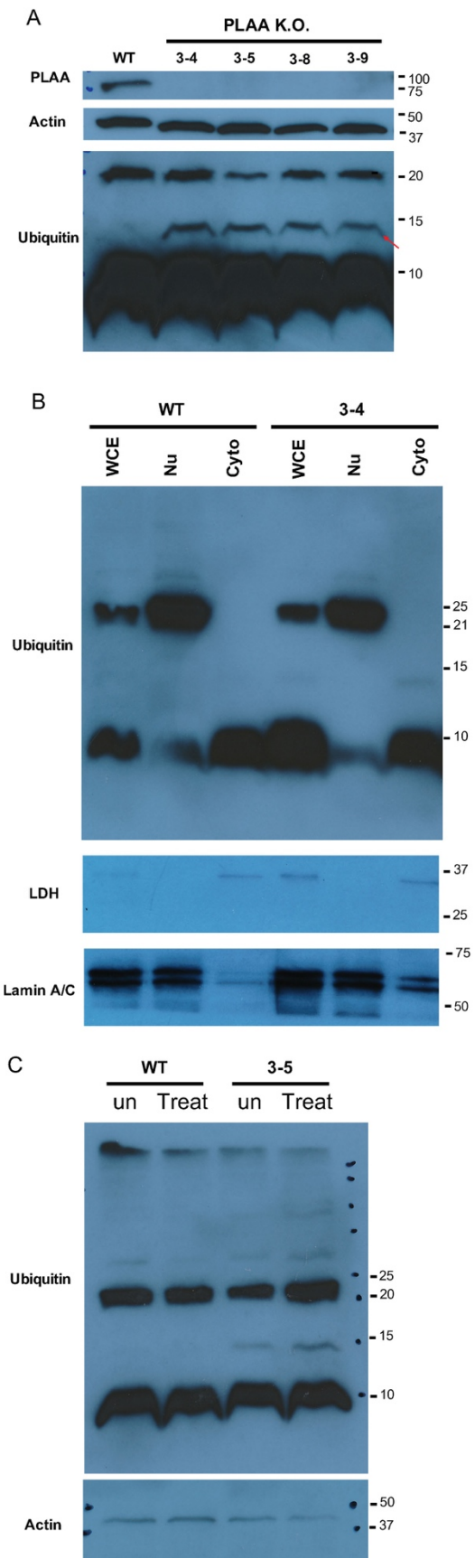


Figure 10: PLAA knockout clones generated using CRISPR/Cas9 and characterization of 15UBL. (A) Western blot analysis of WT and four PLAA KO clones for PLAA, beta-actin, and ubiquitin. These four clones were labeled as 3-4, 3-5, 3-8, and 3-9 because they were different clones selected from plate #3 generated using the same sgRNA targeting Exon1 of PLAA. (B) Subcellular fractionation assay showing 15UBL localization in the cytosol. LDH and Lamin A/C served as loading controls for cytosol and nuclear fractions, respectively. (C) WT and PLAA KO clone 3-5 were treated with 10 μM PR-619 for 24 hours. Whole-cell lysates were subjected to western blot analysis using antibodies against ubiquitin, and beta-actin, as indicated on the left.

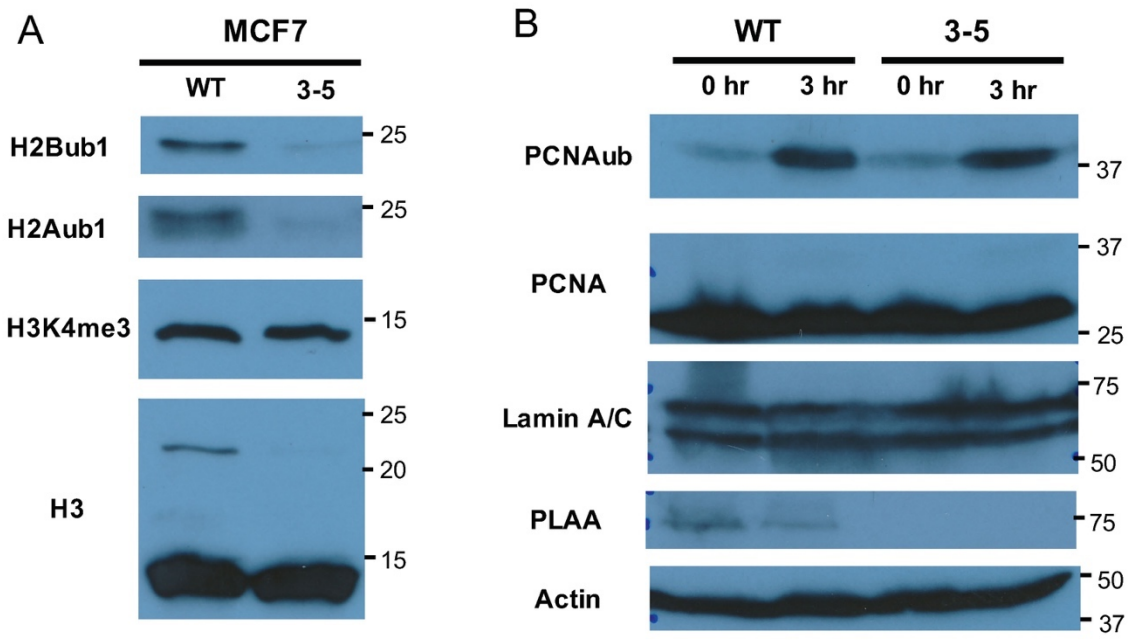


Figure 11: Effects of PLAA depletion on histone and PCNA ubiquitination. Histone extracts (A) and whole-cell lysates (B) from wild type and PLAA KO clone (3-5) were separated by SDS-PAGE and subjected to western blot analysis. B: Both WT and KO cells were treated by UVC light at 30 J/m2 and collected 3 hours post-UVC treatment.

PLAA's function in DNA damage repair

Given that Doa1 facilitates DNA damage response (DDR) by supplying free ubiquitin to the DNA replication and repair machinery (Lis and Romesberg, 2006), I asked whether PLAA is also involved in DDR. I used laser microirradiation followed by immunofluorescence and live-cell imaging to determine the recruitment of PLAA and H2Bub1 to DNA damage sites. Upon UV-induced DNA damage, PLAA was enriched at DNA double-strand breaks (DSBs) as evidenced by its co-localization with γH2AX and RFP-tagged 53BP1. PLAA was enriched, but H2Bub1 was excluded at the DNA damage stripes 15 min post-irradiation (**Figure 12A & 12B**). As PLAA and H2Bub1 antibodies require different fixation methods, I could not perform co-staining using both antibodies. PLAA's recruitment at DSBs was also observed using live-cell imaging. GFP-tagged PLAA was enriched at DSBs indicated by RFP-tagged 53BP1. It's unclear whether PLAA's recruitment to DSBs was dependent on H2Bub1, as PLAA interacts with other DDR-related proteins, including VCP (Acs et al., 2011).

PLAA forms homo-multimer and interacts with VCP

Effector proteins often dimerize or multimerize to create avidity effects, thus enhancing their binding affinities with substrates (Hjerpe et al., 2009). I tested whether PLAA can also form oligomers *in vivo* by co-immunoprecipitation. I co-expressed 3xFlag tagged PLAA and GFP-fused PLAA full-length and four truncated proteins, WD40-PFU (G-W-P), PFU (G-P), PFU-PUL (G-P-P), and PFU deletion (G-PFUΔ). Co-immunoprecipitation results showed that PUL domain is required for PLAA oligomerization while the WD40 and PFU domains are dispensable.

Doa1 interacts with Cdc48 directly via its C-terminal PUL domain. I asked whether PLAA-VCP interaction was conserved in mammalian cells. Consistent with previous studies (Decottignies

et al., 2004, Ghislain et al., 1996, Ogiso et al., 2004, Mullally et al., 2006), PLAA interacts with VCP and PUL domain is required for this interaction (**Figure 13**).

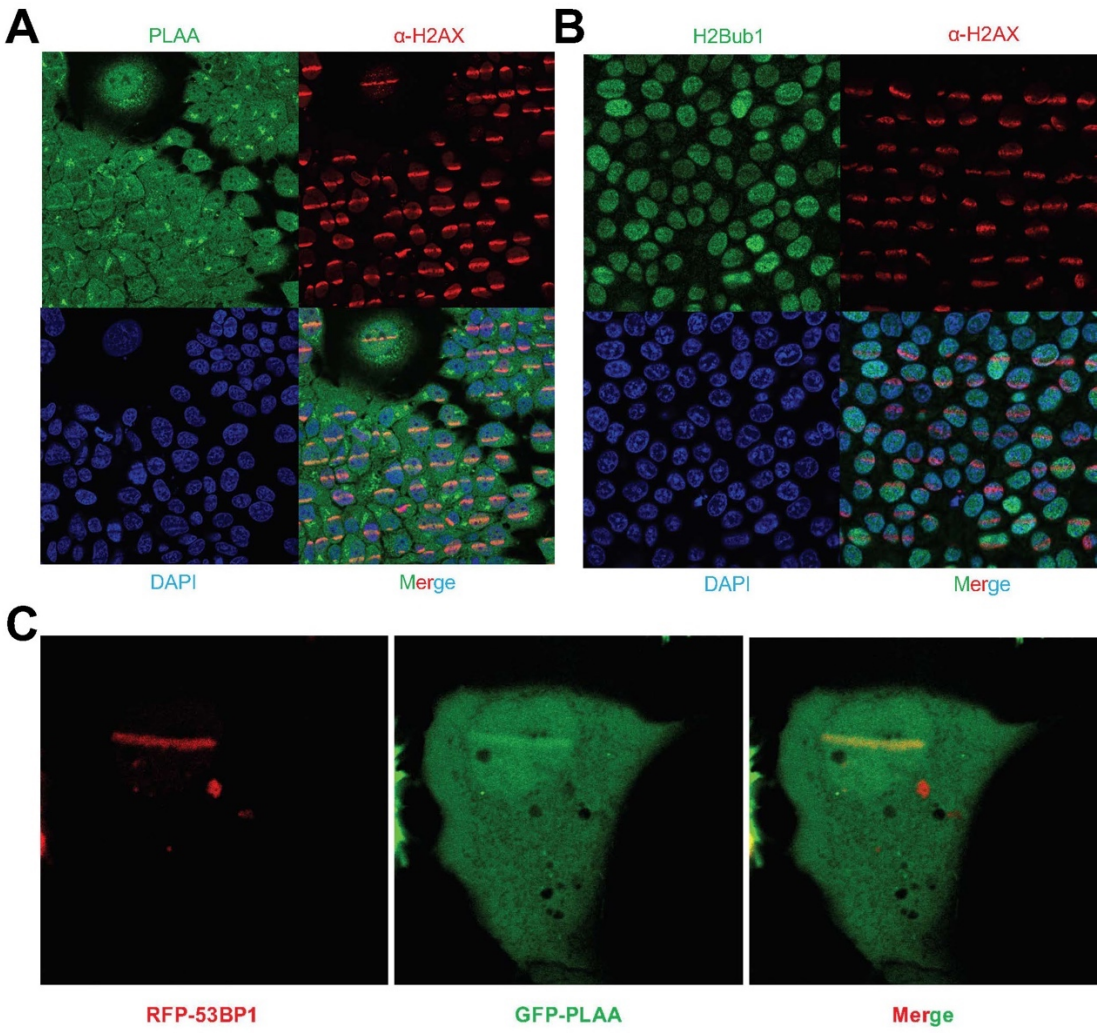


Figure 12. PLAA localizes to, and H2Bub1 is removed at DNA damage sites. (A) PLAA localizes to sites of laser-induced DNA damage in MCF-7 cells. DNA damage was induced by laser microirradiation followed by immunofluorescence (15 min after damage) with indicated antibodies. γH2AX marks the 'laser lines' containing damaged DNA. (B) H2Bub1 is decreased at DNA damage sites in MCF-7 cells. DNA damage was induced as in (A), followed by

immunofluorescence for H2Bub1 and γH2AX. (C) Accumulation of GFP-PLAA and RFP-53BP1 to sites of laser microirradiation shown by live-cell imaging.

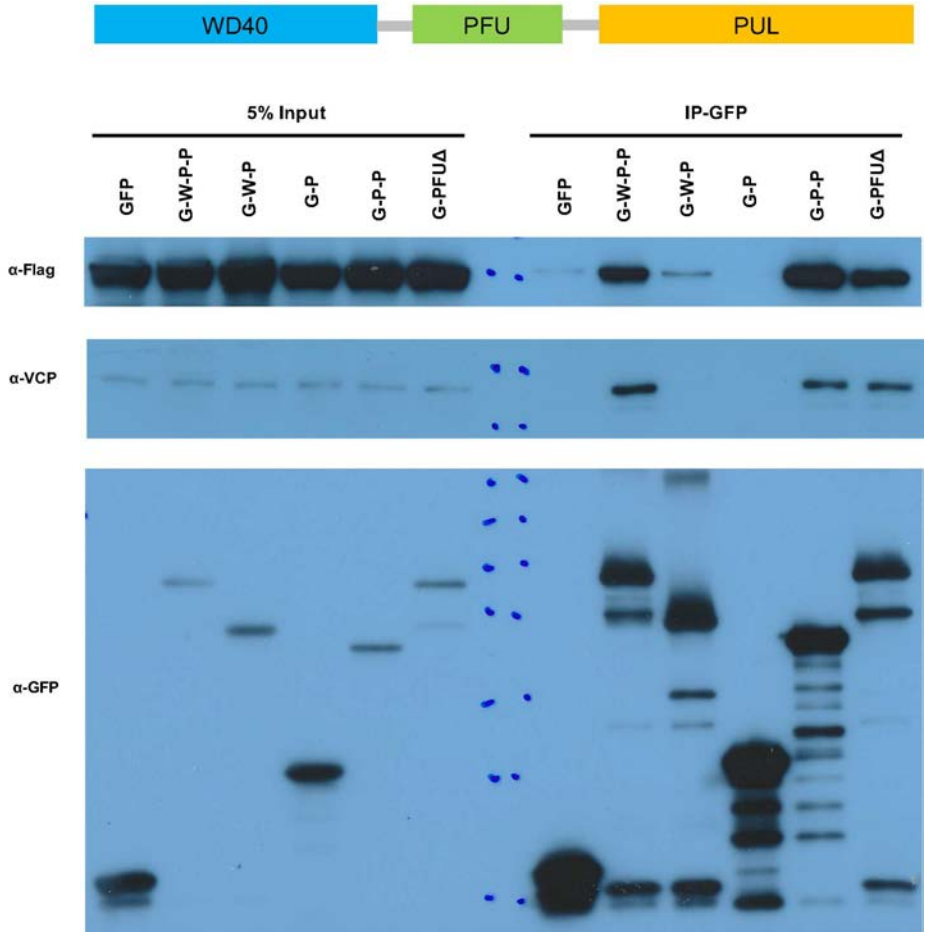


Figure 13. PLAA forms oligomers and interacts with VCP. (A) PLAA has an N-terminal WD40 domain, PFU domain and a C-terminal domain. (B) 3xFlag-PLAA and GFP-tagged full-length and truncated PLAA were co-transfected in 293T cells. G-W-P-P: full-length PLAA, G-W-P: GFP-WD40-PFU, G-P: GFP-PFU, G-P-P: GFP-PFU-PUL, G-PFUΔ: GFP-WD40-PUL. GFP trap was used to coimmunoprecipitate 3xFlag-tagged PLAA and VCP. GFP only was used as a negative control. 5% Input and immunoprecipitates were detected immunochemically using antibodies to Flag, VCP and GFP.

CHAPTER 5: HISTONE H3 N-TERMINAL MIMICRY DRIVES A NOVEL NETWORK OF METHYL-EFFECTOR INTERACTIONS

Part of this chapter is based upon: J. Chen, C. Sagum, J. Zhou, J. Horton J, X. Cheng[3] and M.T. Bedford, Histone H3 N-terminal mimicry drives a novel network of methyl-effector interactions. Under revision.

5.1 INTRODUCTION

This study seeks to address whether there are proteins in the human proteome that have an H3-like N-terminal motif and whether they can interact with known H3K4me3-reader proteins in a methylation-dependent manner, which could dramatically expand the PHD interaction network. H3 N-terminal mimicry proteins (H3TMs) are defined as proteins that start with an iM-Z1-R2-X3-K4 sequence motif. In this motif, iM is the initiator methionine, and Z1 represents one of the seven amino acids that allows for iMet cleavage, and X3 represents any amino acid. First, we searched the database for proteins that harbor this relaxed motif when it is positioned just after the iMet residue, and we identified 48 H3TMs. We chose seven of these H3TMs as candidates for further *in vitro* validation experiments. Next, we used an in-house methyl reader domain microarray, which harbors over 200 methyl effector domains, and identified trimethylation-specific interactions between the seven H3TMs and eight arrayed PHD fingers (and one Tudor domain-containing protein). We validated a sub-set of these interactions by both peptide pull-down of recombinant GST-fusion proteins and their endogenous counterparts from total cell lysates. Using the same pull-down approach, we further show that a number of the unmethylated H3TMs interact with the NURD complex. We also explored the possibility that the K4 site of H3TMs could be post-translationally modified by three known H3K4 methyltransferases: Set7/9, PRDM9 and the MLL1 complex. This study demonstrates that K4 methylation of H3TMs promotes a new interaction network with PHD fingers and blocks their interactions with the repressive NuRD complex.

5.2 RESULTS AND DISCUSSIONS

Bioinformatic search for histone H3 N-terminal mimicry proteins

To identify H3 N-terminal mimicry proteins that undergo initial methionine cleavage and contain RxK motifs, we searched for M-Z-R-X-K motif-containing proteins using the Motif Search database (https://www.genome.jp/tools/motif/MOTIF2.html). Z represents alanine, glycine, serine, cysteine, threonine, proline, or valine, which is a critical component of this motif because these seven amino acids are permissive for initiator methionine cleavage at the P1' position (Sherman et al., 1985, Frottin et al., 2006, Wingfield, 2017a). X represents all 20 possible amino acid residues. We excluded histone H3 variants from the results and identified 257 M-Z-R-X-K motif-containing proteins, which harbored this motif in either the body of the protein or at the protein's N-terminus. About 1/5 of these proteins have the motif restricted to their N-terminus; specifically, 48 of these proteins possessed the motif at their N-termini (**Figure 14A,** lower panel), while the remaining 209 proteins possessed the motif at the internal regions (**Figure 14A,** middle panel). Functional annotation by the Panther database (http://pantherdb.org/) shows that the slice of the pie chart represented by "chromatin-binding protein" (**Figure 14A,** upper panel) does not increase if the M-Z-R-X-K motif is internal. However, if the iM-Z-R-X-K motif was restricted to the N-terminus of proteins, there was a dramatic enrichment of proteins that are functionally annotated as "chromatin-binding protein" (**Figure 14A,** blue wedge). These results indicate that there is a dramatic enrichment of "chromatin-binding protein" class in H3TMs, which indicates that these types of proteins (transcription factors and transcriptional co-regulators) may engage the PHD interaction network.

Phylogenetic analysis of the 48 proteins that harbor iM-Z-R-X-K motifs, identified several families of transcriptional factors with conserved N-terminal sequences: MEF2A/B/C/D, SALL1/2/3/4, BCL11A/B, FOG1/2, and TSHZ1/2/3 (**Figure 14B**). Notably, SALL1(Lauberth

and Rauchman, 2006), FOG1 (Hong et al., 2005, Svensson et al., 2000), and BCL11A/B (Dubuissez et al., 2016b, Moody et al., 2018) are known to interact with the NuRD corepressor complex via their N-termini. We chose one protein from each of the families, MEF2C, BCL11B, and TSHZ1, to perform follow-up studies. VRK1 is a nuclear serine/threonine kinase that phosphorylates several essential proteins including p53/TP53, histone H3, BANF1, and ATF2 (Nichols et al., 2006, Lopez-Borges and Lazo, 2000, Sevilla et al., 2004, Kang et al., 2007). Importantly, VRK1 K4me1 has been identified using mass spectrometry (MS) implying K4 residue is likely a target for methylation *in vivo* (Wu et al., 2015). PCLAF interacts with PCNA to regulate DNA repair during DNA replication (Povlsen et al., 2012, Kais et al., 2011). MAPK8 is another serine/threonine kinase that involves cell proliferation, differentiation, migration, and apoptosis (Kurokawa et al., 2000, FLEMING et al., 2000, Gupta et al., 1996, Dai et al., 2000, Miotto and Struhl, 2011, Tomlinson et al., 2010). HIVEP1 is a zinc-finger protein that binds to the HIV-1 enhancer to activate HIV-1 gene expression (Maekawa et al., 1989). Thus, we selected seven H3TMs candidates for further evaluation, because of their distribution across the phylogenetic tree and their roles in transcriptional regulation: VRK1, BCL11B, TSHZ1, MEF2C, PCLAF, MAPK8, and HIVEP1 (**Figure 14C**). Then, we screened for potential novel interactions between these candidates and PHD finger proteins.

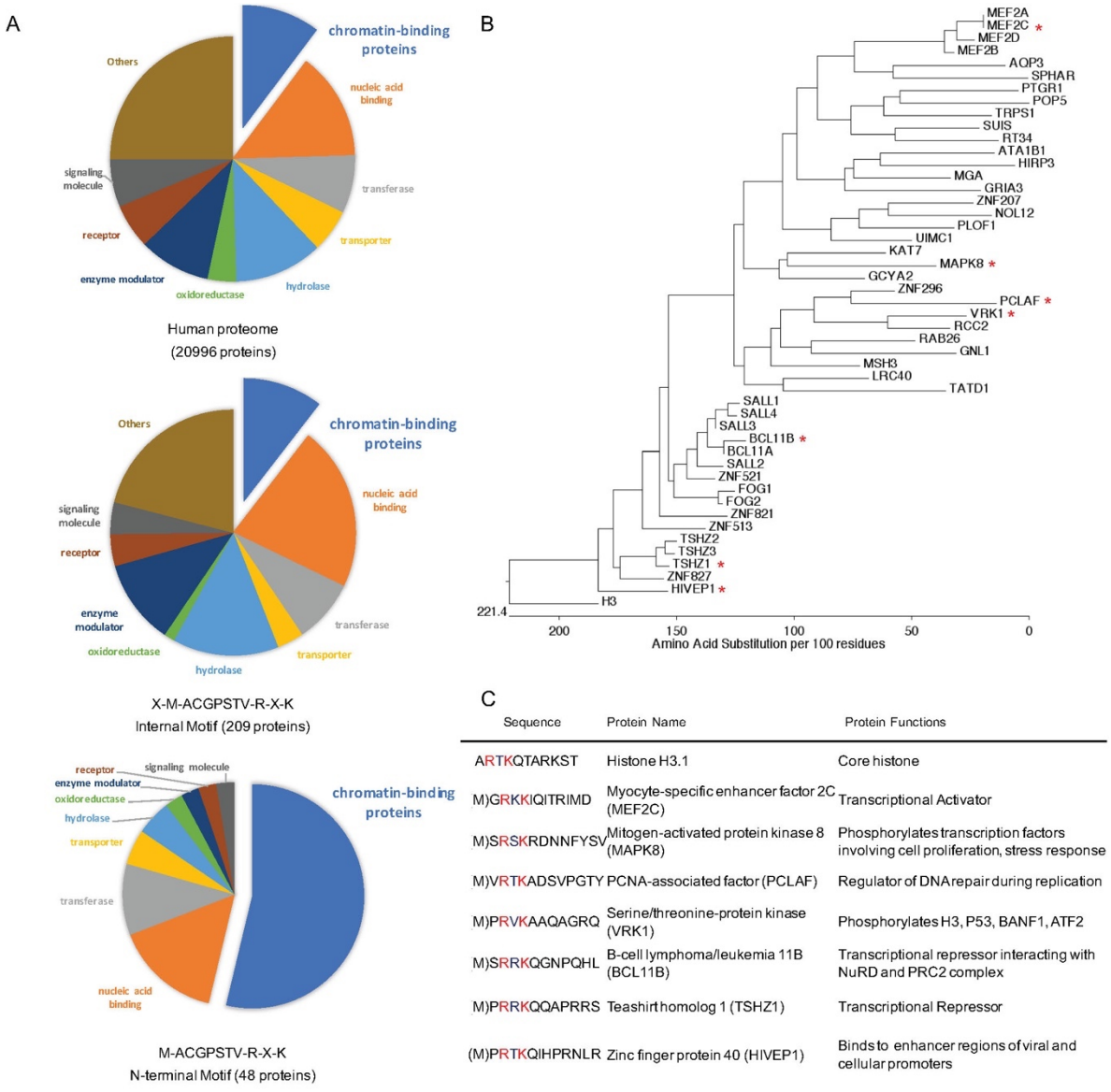


Figure 14. Identification of H3 N-Terminal Mimicry Proteins. A) PANTHER Protein Class annotation of 20996 proteins in the human proteome, 209 internal motif-containing proteins, and 48 N-terminal motif-containing proteins (excluding histone H3 and variants). B). Phylogenetic analysis of the N-terminal sequences of 48 H3TMs and histone H3. Red asterisks indicate seven H3TMs candidate for follow-up study and their N-terminal sequences and functions are summarized in C).

Identification of H3TMs interactions with methyl-lysine effector domains

To identify potential interactions between the 7 candidate proteins and methyl-reading domains, we synthesized N-terminal peptides (2-13 amino acids excluding the initial methionine) with and without tri-methylation at the K4 position. We used these biotinylated peptides as probes to interrogate an in-house methyl-reader microarray that harbors GST fusions of known and predicted binders of methyl-lysine and methyl-arginine marked motifs (Chen et al., 2020). The current array contains 108 PHD, 40 Tudor, and 31 Chromo domains. As positive and negative controls, the histone H3K4me3 peptide interacted with known H3K4me3-interacting PHD finger domains and Tudor domains, while the unmodified peptide did not (**Figure 15**, top left panel). Interestingly, each of the seven peptide probes that harbored the K4me3 mark (**Figure 14C**) displayed a unique interaction profile with known H3K4me3-interacting PHD finger domains (Jain et al., 2020), and many of these peptides were also recognized by the Tudor domain-containing protein SPIN1, which is a well characterized H3K4me3 reader (Yang et al., 2012). Notably, compared to the other probes, VRK1 and BCL11B had a greater number of interacting reader domains. When the K4 position was tri-methylated, both VRK1 and BCL11B interacted with PHF2, TAF3, and KDM7A PHD fingers; however, VRK1 selectively interacted with DIDO1, ING1/2, and MLL5 PHD fingers, while BCL11B selectively interacted with ING3 and PHF23 PHD fingers. The TSHZ1 K4me3 peptide bound the PHD fingers of TAF3 and ING3, and PCLAF only interacted with KDM7A. Also, as six of the seven methylated peptide probes interacted well with the arrayed SPIN1 Tudor domains, we included this domain as a positive control in the downstream validation experiments.

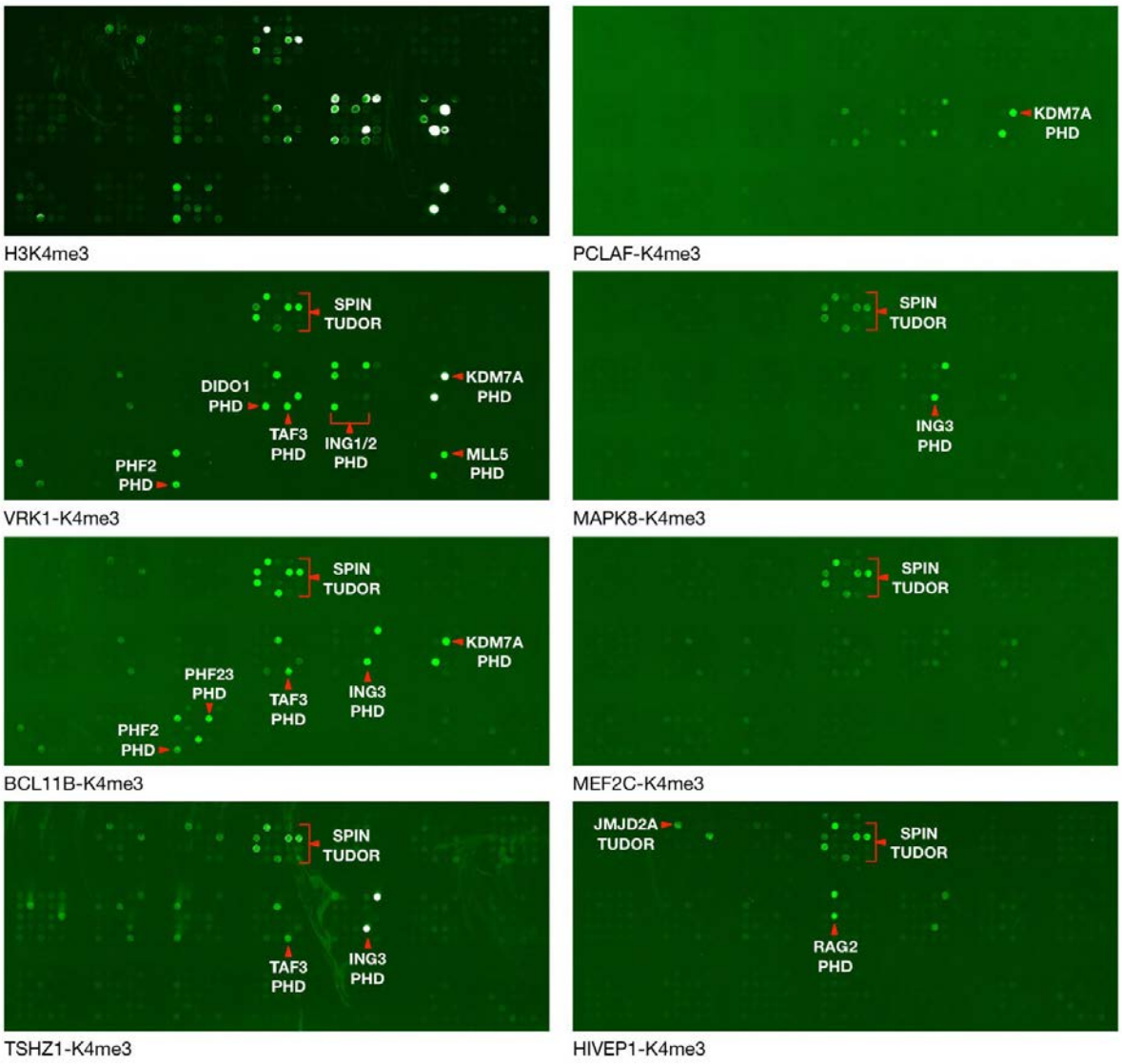


Figure 15. Identification of novel methylation-dependent interactions between H3TMs and PHD fingers using a methyl reader domain microarray. N-terminal peptide probes without (not shown) and with K4me3 for 7 H3TMs and histone H3 were probed onto our methyl-reader microarray. Duplicates of each protein domain were used to facilitate identification and reproducibility. Only methylation-specific interactions were selected for downstream validation.

Validation of novel interactions using *in vitro* interaction assays

We used a peptide pull-down assay to confirm a small subset of the interactions we observed on the protein domain microarrays, and to validate the methyl-dependent nature of these bindings. GST fusions of PHD domains from PHF2, ING2 and ING3, and the Tudor domain of SPIN1 were used to perform pull-downs with biotinylated VRK1, BCL11B, TSHZ1 and histone H3 peptide pairs, in either methylated or unmethylated forms. The PHF2 PHD finger strongly interacted with the methylated peptides derived from VRK1, BCL11B, and H3; whereas it interacted more weakly with the methylated form of TSHZ1. The ING2 PHD finger only bound with the methylated VRK1 and H3 peptides. ING3 showed methylation-dependent interactions with TSHZ1 and H3, but could also interact with VRK1 in a methyl-independent fashion. SPIN1 interacted with VRK1, BCL11B, TSHZ1, and H3 peptides in their tri-methylated forms (**Figure 3A**).

Next, we asked if we could use the same peptide sets to pull down the full-length methyl-effector proteins from whole-cell lysates. We performed these experiments because it is possible that the isolated recombinant protein domains may behave differently than the full-length endogenous counterparts. To facilitate this experiment, we were able to identify antibodies that could recognize endogenous PHF2, ING2 and SPIN1, but we were unable to find a good ING3 antibody. The whole-cell lysates pull-downs using the VRK1 peptide sets, enriched for endogenous PHF2, ING2 and SPIN1, when methylated. The BCL11B methyl-peptide enriched for endogenous SPIN1, and very weakly endogenous PHF2. The TSHZ1 methyl-peptide only enriched for endogenous SPIN1 (**Figure 16B**).

A number of the H3MTs listed in the phylogenetic tree (**Figure 1C**) harbor the ability to bind the NuRD repressor complex, through their N-termini. For example, SALL1 (Lauberth and Rauchman, 2006), FOG1 (Hong et al., 2005, Svensson et al., 2000), and BCL11A/B

(Dubuissez et al., 2016b, Moody et al., 2018) all possess this property, and it is thought that methylation at their K4 position will block this interaction just as it does for the interaction between NuRD and Histone H3 (Zegerman et al., 2002). It has been shown that the MTA1 is a core subunit of the NurD complex, and the C-terminus of MTA1 (454–715) recognizes the N-terminus of BCL11B (Wu et al., 2013b). We next determined whether K4me3 has the ability to disrupt the interactions between the NuRD complex and VRK1, BCL11B, and TSHZ1. To do this, we again use the peptide sets to perform pull-downs from whole-cell lysates, and we probed these pull-downs with antibodies against MTA1 and RBAP46/48 (another component of the NuRD complex). Indeed, as expected, we find that BCL11B and histone H3 peptides engage NuRD when unmethylated, and so does TSHZ1, which has not been reported before. The unmethylated VRK1 peptide does not have the ability to recruit the NuRD complex (**Figure 16B**). Thus, just like histone H3, a number of H3MTs can switch from recruiters of activator complexes (in their methylated state) to recruiters of the NuRD repressor complex (in their unmethylated state).

In our pull-down assays, performed from whole-cell lysates, we noticed that substantially more PHF2 was enriched using the VRK1-K4me3 peptide as compared to the H3K4me3 peptide (**Figure 16B**). PHF2 PHD domain has been reported to have a binding affinity (Kd) of 230 nM for H3K4me3 (Horton et al., 2011), and we performed isothermal titration calorimetry (ITC) to determine the binding affinity of the PHF PHD finger for the methylate N-terminus of VRK1. As suggested by the pull-downs, we observed an almost three-fold stronger binding of PHF2 for VRK1 than for H3K4me3 (87 nM vs 230 nM). The binding affinity of PHF2 for BCL11B was much weaker at 7.1 uM (**Figure 16C**). Thus, there are a myriad of interactions that can potentially occur on H3TMs, either in their methylated or unmethylated states, and the high binding affinity between the VRK1 K4me3 peptide and the PHF2 PHD domain suggests a

competition model where PHF2 might be recruited to chromatin by H3K4me3 and then handed off to VRK1-K4me3.

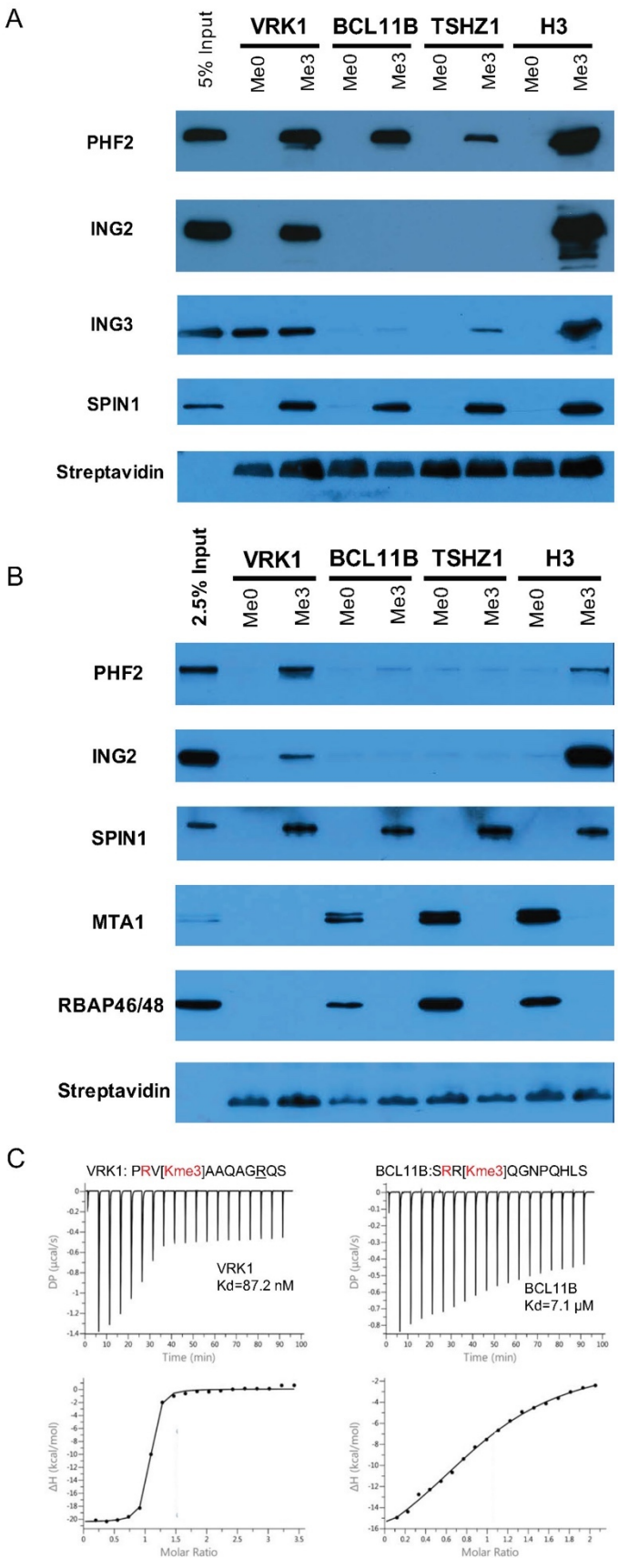


Figure 16. Validation of novel interactions using in vitro interaction assays. A) Peptides pull down GST-tagged PHD fingers. Unmodified and K4me3 peptide probes for VRK1, BCL11B, TSHZ1, and H3 were used to pull down GST-tagged PHF2 PHD, ING2 PHD, ING3 PHD, and SPIN1 TUDOR. B). Peptides pull down endogenous PHD fingers and NuRD complex from 293T whole cell lysates. Peptides in (A) were used to pull down PHF2, ING2, ING3, SPIN1 and MTA1 and RBAP46/48. Immunoblot using streptavidin-HRP was used to ensure equal loading of all biotin-tagged peptide probes. C). Isothermal Titration Calorimetry (ITC) measurement of binding of PHF2(1-451) to tri-methylated VRK1 (1-12) K4me3 peptides. Binding constant was calculated by fitting the data to one-site binding model equation using the ITC data-analysis module of Origin 7.0 (OriginLab Corporation).

The K4 residue of three H3TMs can be methylated by H3K4 methyltransferases *in vitro*

Importantly, VRK1-K4me1 has been identified using mass spectrometry (MS) implying K4 residue is likely a target for methylation *in vivo* (Wu et al., 2015). In addition, three independent MS studies performed by Cell Signaling Technology (CST), publicly available through the PhosphoSitePlus website (www.phosphosite.org), have also identified mono-methylation at the VRK1-K4 position. All four of these reported studies focused on enriching for Kme1 peptides, either using methyl-specific Kme1 antibodies (CST) or using a unique chemical derivation approach followed by antibody enrichment, which was used by Yingming Zhao's group (Wu et al., 2015). These independent studies suggest that VRK1 is indeed targeted for methylation at its K4 position, just like histone H4, and that higher degrees of methylation (me2/3) may occur.

To experimentally test if H3TMs have the potential of being methylated by enzymes that modify the H3K4 site, we set up a number of *in vitro* methylation reactions. Three different lysine methyltransferases that have been reported to deposit methyl marks at the histone H3K4 position were used: Set7/9 (Wang et al., 2001), MLL1 complex (Dou et al., 2006, Song and Kingston, 2008) and PRDM9 (Wu et al., 2013a). We performed *in vitro* methylation using the same set of peptide pairs used in previous experiments. For each set of *in vitro* methylation reaction, enzyme and ^{3}H-labeled Sam served as negative controls, and H3K4 peptides served as positive controls. We showed that Set7/9 could methylate unmodified peptides from all three candidate proteins VRK1, BCL11B and TSHZ1 but could not methylate the K4me3 peptides, indicating that methylation occurs at the K4 sites (**Figure 17**). MLL1 and PRDM9 can only methylate VRK1 with moderate activity, compared to H3 peptide, suggesting that VRK1 is a weaker substrate than histone H3 for MLL1 and PRDM9. These data support the hypothesis that at least a subset of H3TMs can be targeted by both the readers and writers of histone H3K4me3.

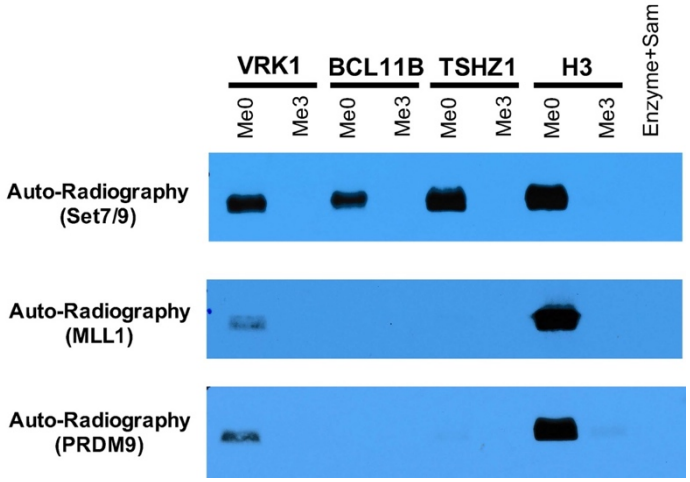


Figure 17. Three H3TMs can be methylated at Lys4 by H3K4 methyltransferases. Unmodified and K4me3 peptide probes for VRK1, BCL11B, TSHZ1, and H3 were incubated with three H3K4-specific methyltransferases, Set7/9 (upper), MLL1 (middle) and PRDM9 (lower). Enzymes and ^{3}H-labeled Sam were used as a negative control (most right line). Immunoblot using streptavidin-HRP was used to ensure equal loading of all biotin-tagged peptide probes.

Generation of C-terminal GFP tagged H3TM Fusion Proteins

To facilitate the characterization of 6 H3TM candidates *in vivo*, I designed a set of fusion proteins consisting of the H3TMs linked to a C-terminal GFP (**Figure 18A**). Besides serving as a purification tag, the C-terminal GFP tag keeps the N-termini of H3TMs intact and available for modifications and protein interactions.

All six fusion proteins showed good expression levels in 293T cells at the expected positions (**Figure 18B**). GFP fusion proteins can be efficiently purified from whole-cell lysates using a GFP-trap® (ChromoTek, Germany), which utilizes beads that are covalently attached to a nanobody that only recognizes GFP, thus removing the confounding issues of heavy or light chain contamination that is associated with traditional antibody-based enrichment approaches (**Figure 18C**).

Subcellular localization of GFP-tagged H3TMs

I examined the subcellular localization of six H3TM-GFP fusions in MCF-7 cells (**Figure 19**). VRK1, BCL11B, and MEF2C are exclusively in the nucleus. PCLAF and TSHZ1 displayed nuclear enrichment, while MAPK8 are homogeneously localized in both the nucleus and cytoplasm. Interestingly, BCL11B, MEF2C, and PCLAF exhibited striking punctate patterns in the nucleus. It remains unknown how their speckled patterns relate to or affect their nuclear functions.

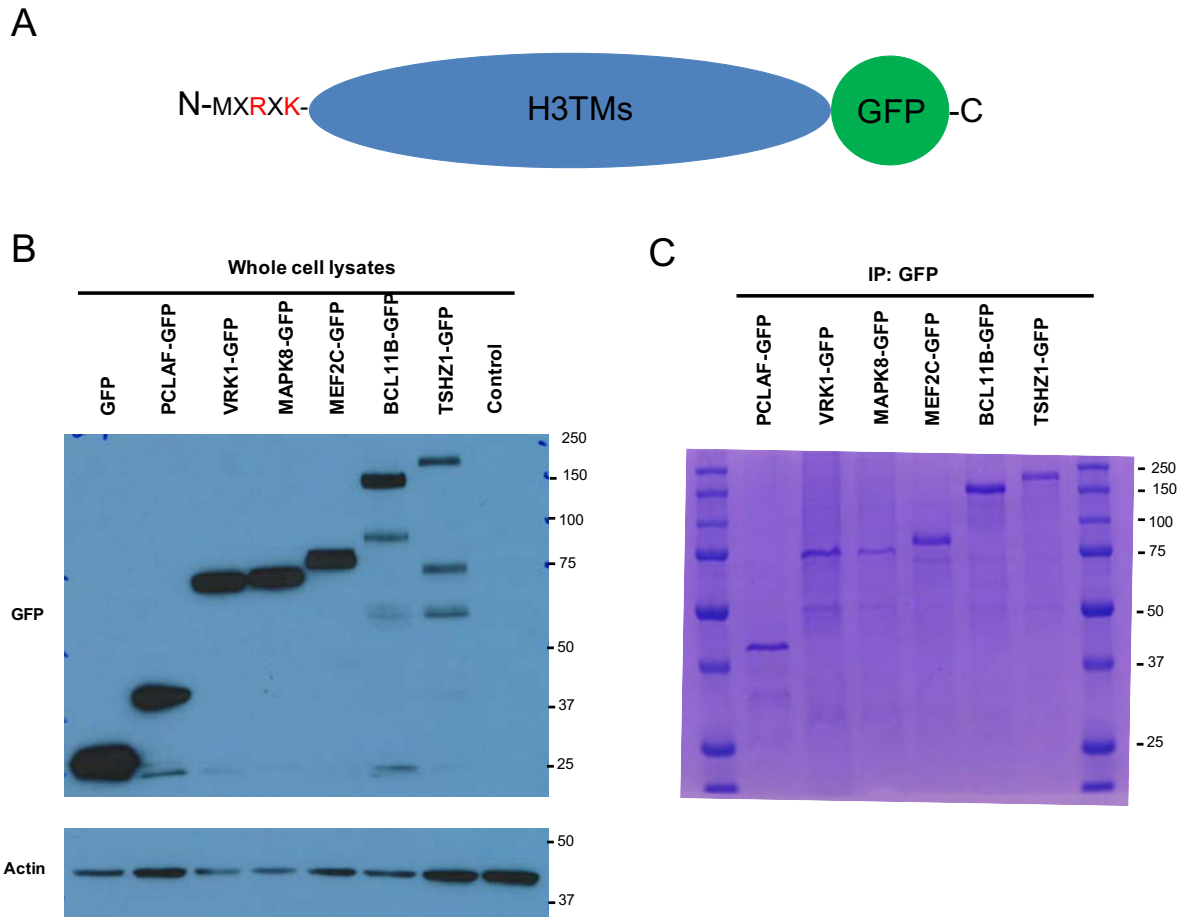


Figure 18: Design and generation of H3TM-GFP fusion proteins. (A) The cDNAs of six H3TMs were synthesized with codon optimization and cloned into the pEGFP-N1 vector by Biomatik (Ontario, Canada). (B) Overexpression of H3TM-GFP fusion proteins and GFP only vector in 293T cell lines. Whole-cell lysates were analyzed using western blot against GFP and beta-actin. Control: an untransfected control. (C) Ectopically expressed H3TM-GFP fusion proteins were immunoprecipitated using the GFP-trap with great purity.

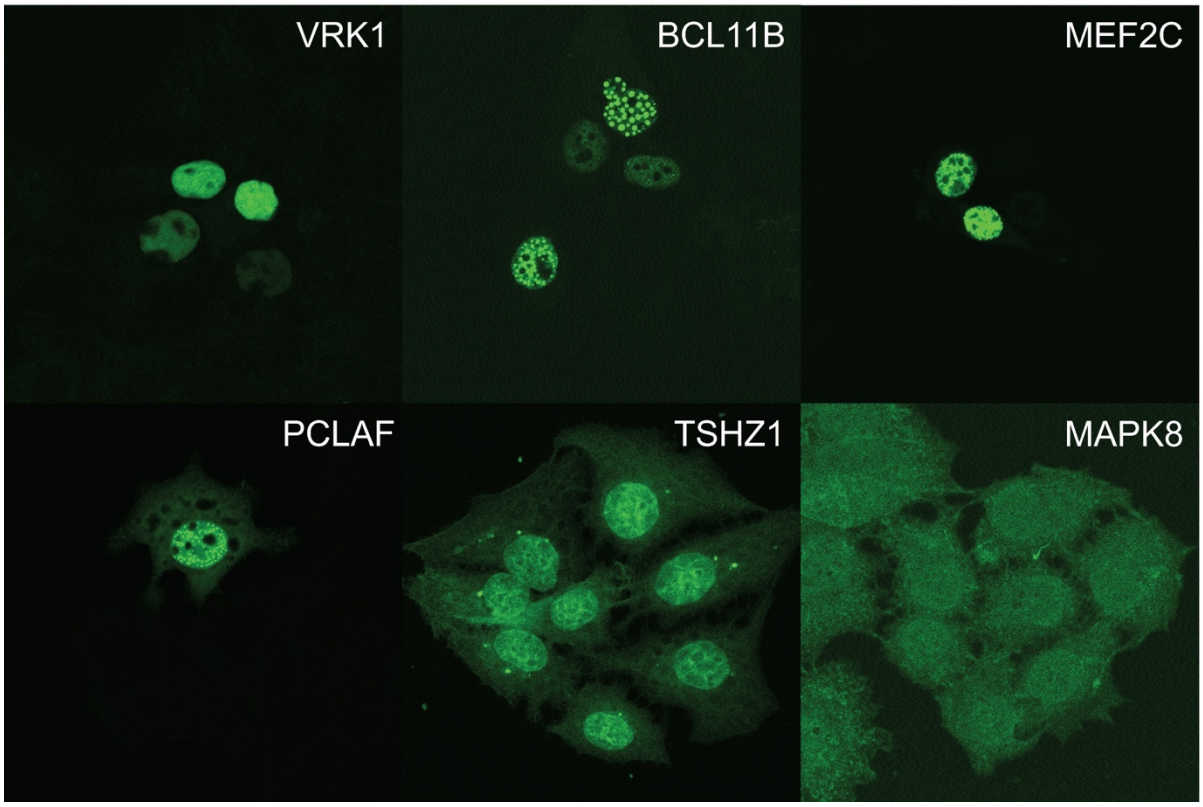


Figure 19. Subcellular localization of H3TMs in MCF-7 cells. MCF-7 cells were transiently transfected with the six H3TM-GFP fusions, fixed, and stained for F-actin and DAPI. Immunofluorescence was detected using a Carl Zeiss LSM880 microscope equipped with an Aryscan confocal microscope (Zeiss, Jena, Germany)

Mapping of H3TM K4me3 using LC-MS

I overexpressed H3TM-GFP fusion proteins in 293T cells and purified them using the GFP-Trap kit. The enriched samples were separated by SDS-PAGE and followed by Coomassie staining (**Figure 18C**). The dominant gel bands corresponding to the full-length GFP fusions were cut and sent to the Proteomics Facility at UT Austin for digestion and LC-MS analysis. Tryptic digestion only covered the N-terminal sequences of BCL11B and PCLAF, and no K4 methylation was identified in these two cases. For the remaining four candidates, tryptic digestion was not able to generate peptides that covered the K4 position, likely because these regions are rich in arginine and lysine residues thus cleaved into very small fragments by trypsin. Digestion using GluC did not cover any N-terminal sequences of all six proteins either (**Table 5**).

Failure to identify the N-terminal sequences might result from the "masking" effects caused by other highly abundant peptides. To overcome this obstacle, I cloned a TEV cleavage site into the N-terminus of VRK1, between Q11 and S12 residues. Upon TEV cleavage, the supernatant was supposed to contain the N-terminus of VRK1-fused with TEV site as the most dominant peptide, which was sent for LC-MS analysis. However, the most abundant peptide (peak intensity: 4E8) started with A5 residue. The second most abundant peptide started with K4 but had an intensity that was 300 folds less (peak intensity: 1.3E6), suggesting that VRK1 N-terminus was subjective to endogenous cleavage that limited its capability to be mapped by LC-MS analysis (**Table 6**). Thus, using these mass spec approaches, we were unable confirm K4 methylation in our H3TMs.

Table 6. Identification of N-terminal sequences of H3TMs using LC-MS

	N-terminal Sequence Coverage by Different Digestion Enzyme		
H3TM-GFP Fusion	Trypsin	GluC	TEV
VRK1	No	No	
BCL11B	Yes, No K4me1/2/3	No	
MEF2C	No	No	
PCLAF	Yes, No K4me1/2/3	No	
MAPK8	No	No	
TSHZ1	No	No	
VRK1-TEV			AAQAGRQ*ENLYFQ (4E8)* VKAAQAGRQ*ENLYFQ (1.3E6)* *(*M)PRVKAAQAGRQ*ENLYFQ* *(Not identified)*

Development and initial characterization of K4me3-Specific Antibodies

As LC-MS analysis didn't find evidence of K4me3, I developed K4me3- and N-terminus-specific antibodies for each of the four candidates, VRK1, BCL11B, MEF2C, and TSHZ1. The K4 sites of these four proteins are subject to mono-methylation, acetylation, and ubiquitination (Wu et al., 2015, Angelelli et al., 2008, Dubuissez et al., 2016a, Schwer et al., 2009), similar to the H3K4 site, suggesting that they are potentially modified by tri-methylation. Antigens contain the first 12 amino acids without iMet and with K4me3. Two polyclonal antibodies were generated for each candidate through affinity purification, which recognized the K4me3 motif and N-terminal sequence, respectively. I first tested the antibody specificity using a dot blot assay. K4me3-specific antibodies only detected K4me3 peptide probes, while N-term-specific antibodies recognized both modified and unmodified peptide probes, indicating excellent specificity at the peptide levels (**Figure 20A**).

I also tested whether these antibodies could recognize the H3TM-GFP fusion proteins. I generated K4A mutants for all four candidates as negative controls. I overexpressed both WT and K4A proteins in 293T cell lines and purified the fusion proteins using GFP-trap, followed by western blot analysis. As expected, all four N-term antibodies detected both WT and K4A signals at the calculated molecular weights. However, K4me3 antibodies for VRK1 and TSHZ1 recognized both WT and K4A mutant proteins, while BCL11B and MEF2C K4me3 antibodies didn't detect signals of the expected molecular weights.

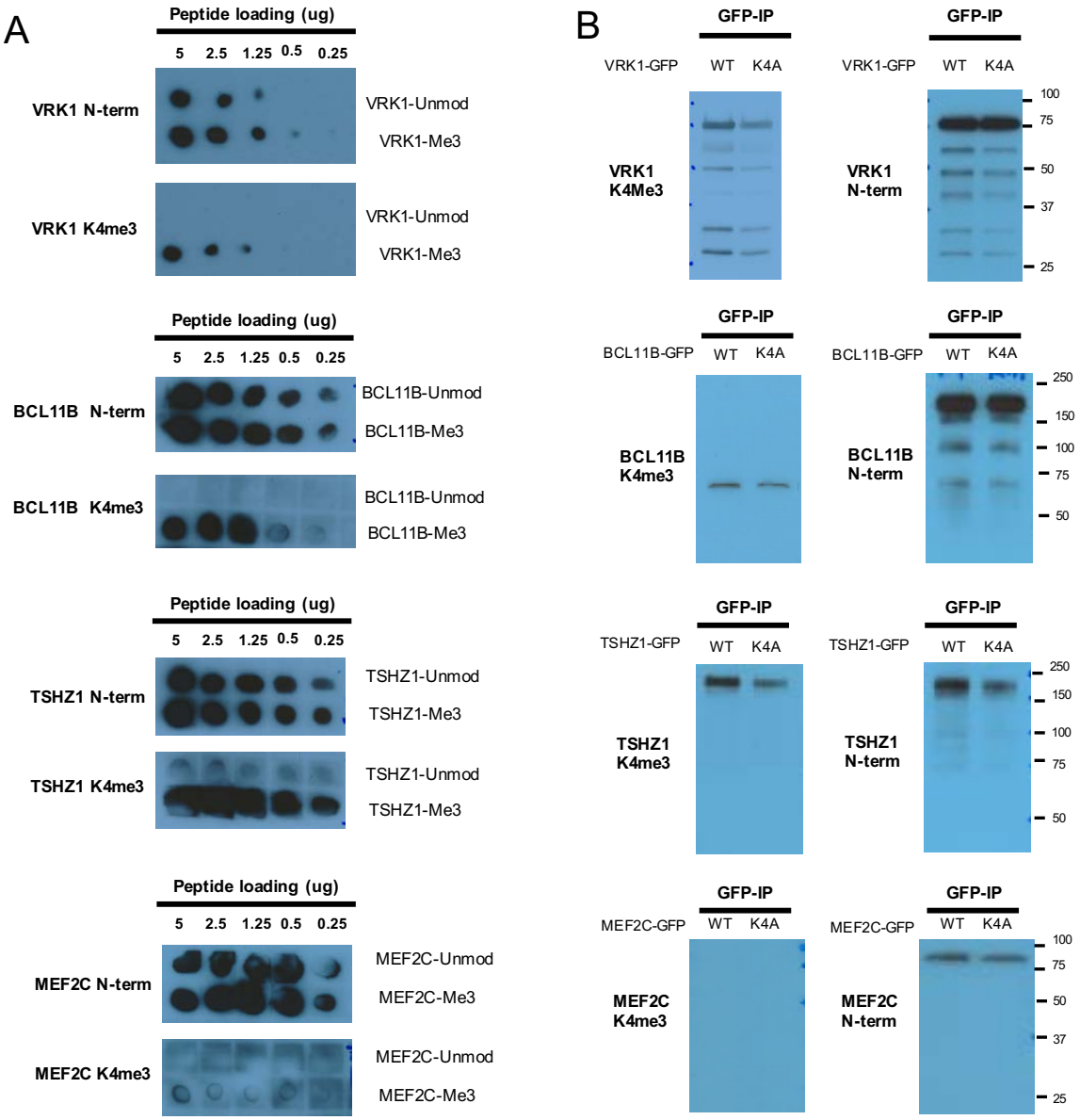


Figure 20. Development and validation of K4me3-specific antibodies. (A) Validation of antibody specificity using dot blot assays. K4me3 and unmodified peptide probes were blotted onto the membrane with decreasing concentrations, 5, 2.5, 1.25, 0.5, and 0.25 μg. Duplicate membranes were prepared and probed with the indicated antibodies. (B) Overexpression of VRK1-GFP, BCL11B-GFP, TSHZ1-GFP, and MEF2C-GFP proteins (WT vs. K4A) in 293T cells. H3TM-GFP proteins were immunoprecipitated using GFP-Trap and analyzed by western blot with the indicated antibodies.

Characterization of VRK1 K4me3-Specific Antibody

Results from **Figure 20** suggested that the VRK1 K4me3 antibody had non-specific binding activity towards unmodified VRK1 full-length protein. I then tested the specificity of the VRK1 K4me3 antibody using the VRK1 unmodified peptides and H3K4me3/unmodified peptides. Both VRK1 K4me3 and H3K4me3 peptides strongly interacted with the tested antibody, though VRK1 K4me3 was used as the immunogen. VRK1 K4me3 antibody displayed pan-me3 binding affinity. Unmodified VRK1 and H3 peptides pulled down a small amount of the tested antibody compared to Kme3 peptides (**Figure 21A**), suggesting that this polyclonal antibody had a non-specific binding affinity towards the unmodified VRK1 and H3 N-terminal motifs. Along with the observations from **Figure 20A**, the antibody's inadequate specificity prevented its application in qualitative validation of the VRK1 K4me3 mark *in vivo*.

To improve the specificity of the VRK1 K4me3 antibody, I developed two negative depletion methods. Method 1 was based on dot blot assay. VRK1 K4me3 antibody was incubated overnight with a membrane containing 60 dots of VRK1 unmodified peptide (5 μg peptide per dot) before being used for western blot analysis. Method 2 involved preincubation of 2 μL VRK1 K4me3 antibody with 10 μg VRK1 unmodified peptides in a total volume of 20 μL for 1 hour at room temperature. Without depletion treatment, VRK1 K4me3 recognized both WT and K4A mutant proteins. In contrast, antibodies treated with two depletion methods could not detect either WT or K4A mutant proteins after the same exposure times (**Figure 21B**). These results suggested GFP-VRK1 was not trimethylated at the K4 site under the current experimental conditions. The K4me3 antibody was pretreated using method 2 before western blot analysis for all subsequent experiments.

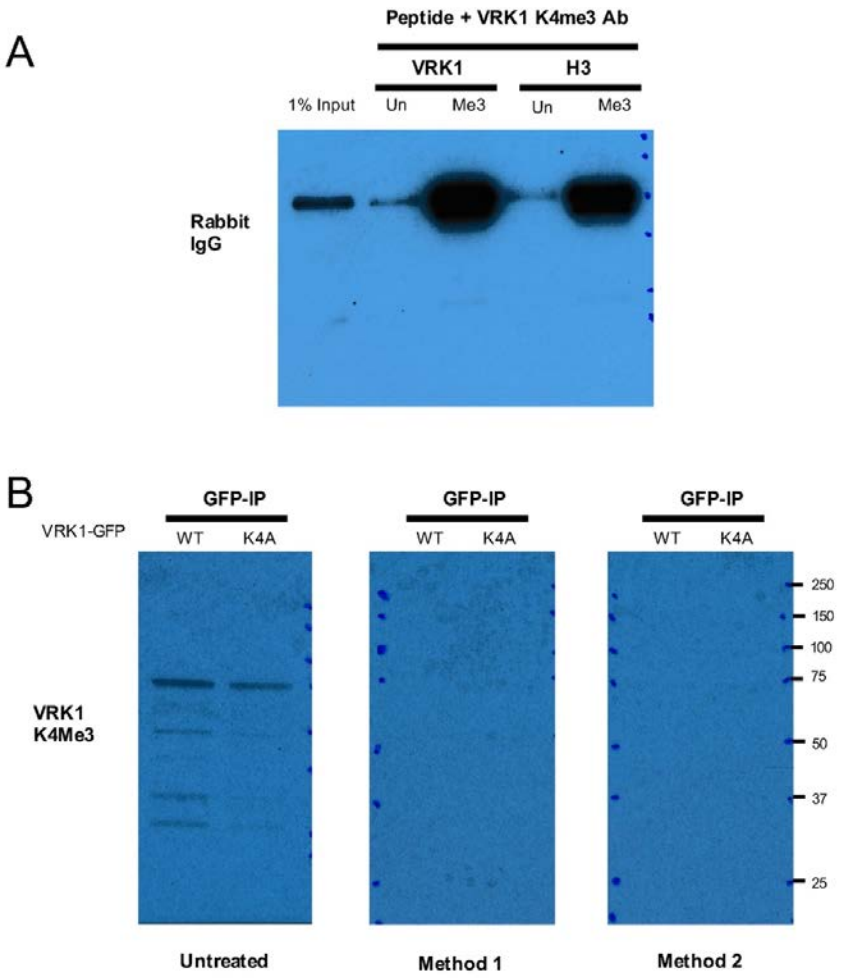


Figure 21 Characterization of VRK1 K4me3 antibody and Negative Depletion Treatment

(A) VRK1 and Histone H3 peptides (unmodified and K4me3) were used to pull down the VRK1 K4me3 antibody. The samples were separated by SDS-PAGE and analyzed using western blot with the anti-Rabbit secondary antibody. (B) Comparison of two negative depletion methods. VRK1-GFP (WT and K4A) fusion proteins were overexpressed in 293T cells and purified using GFP-Trap. Three membranes were probed with untreated and treated antibodies. Left: untreated antibody. Middle: method 1: 2 μL antibody was pretreated with a membrane blotted with 60 VRK1 unmodified peptide dots (5 μg peptide per dot) overnight at 4°C. Right: method 2: 2 μL antibody was incubated with 20 μg of unmodified VRK1 peptides for 1 hour at room temperature.

After improving the VRK1 K4me3 antibody specificity by negative depletion, I further characterized this antibody using VRK1 knockdown and overexpression of VRK1-GFP proteins. VRK1 knockdown and overexpression were confirmed by both the VRK1 (Sigma) and VRK1 N-term antibodies **(Figure 22A)**. VRK1 (Sigma) antibody recognizes the C-terminus of VRK1. VRK1 K4me3 detected multiple bands across four samples, and these signals didn't change or shift accordingly to those detected by the VRK1 N-term antibody **(Figure 22A**). I also purified endogenous VRK1 from control and VRK1 KD cells using the VRK1 (Sigma) antibody. Similarly, VRK1 K4me3 signals didn't decrease in VRK1 KD cells compared to control cells, suggesting that VRK1 K4me3 signals were false-positive signals **(Figure 22B**).

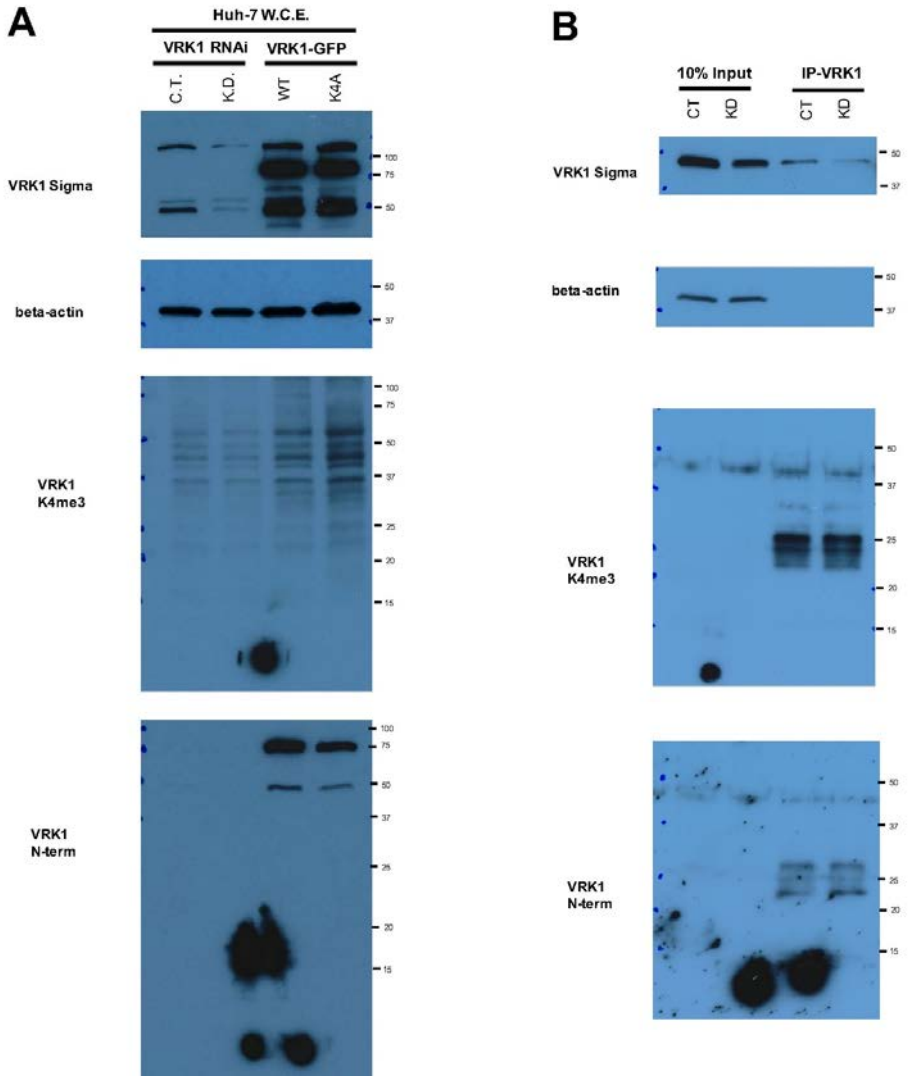


Figure 22. Characterization of VRK1 K4me3 antibody using VRK1 knockdown and VRK1-GFP overexpression. (A) Huh-7 cells were transiently transfected with control siRNA, VRK1 siRNA, WT VRK1-GFP and K4A-GFP plasmids. Whole-cell lysates were separated by SDS-PAGE and analyzed by western blot against VRK1 (Sigma), beta-actin, VRK1 K4me3 (after negative depletion) and VRK1 N-terminus. (B) Huh-7 cells were transfected with control siRNA and VRK1 siRNA. Endogenous VRK1 was immunoprecipitated with the anti-VRK1 antibody (Sigma) and analyzed by western blot using the same antibody set. At the bottom of membranes in (A) and (B), 2.5 ug of K4me3 (left) and unmodified (right) VRK1 peptides were blotted on the membrane as positive controls.

Detection of VRK1 K4me3 and VRK1-PHF2 interactions under various circumstances

VRK1 has the highest expression levels in the G2/M phase and phosphorylates histone H3 at Thr3 and Ser10 (Kang et al., 2007). I asked whether mitotic arrest could upregulate VRK1 K4me3 *in vivo* using the Nocodazole treatment. Western blot analysis of input and immunoprecipitated samples displayed comparable VRK1 levels and VRK1 K4me3 signals (**Figure 23A**), indicating that VRK1 K4me3 cannot be induced by mitotic arrest.

Set7/9, a mono- and di-methyltransferase, can methylate the VRK1 K4 site (Hamidi et al., 2018). I tested whether overexpression of Set7/9 could stimulate K4me3 signals and switch on the VRK1-PHF2 interactions by activating the VRK1 K4 site for potential tri-methyltransferase by depositing mono- and dimethyl marks. PHF2 was not co-immunoprecipitated by VRK1 and VRK1 K4me3 signals didn't change across three cell lines (**Figure 23B**). RPL29K5me2 was not significantly upregulated in WT Set7/9 overexpressed cells, possibly resulting from high expression levels of endogenous Set7/9 in 293T cells (Hamidi et al., 2018). All the above results suggested that Set7/9 overexpression could not induce VRK1 K4me3 signals or turn on VRK1-PHF2 interactions.

I also tested whether PHF2 could be co-immunoprecipitated by the WT VRK1-GFP fusion protein. I transiently transfected 293T cells with WT and K4A mutant VRK1-GFP and purified the fusion proteins using GFP-Trap and detected with the anti-PHF2 antibody. Consistent with **Figure 23B**, PHF2 was not co-immunoprecipitated with WT or K4A mutant VRK1-GFP fusions, suggesting no VRK1 K4me3 under the current experimental conditions.

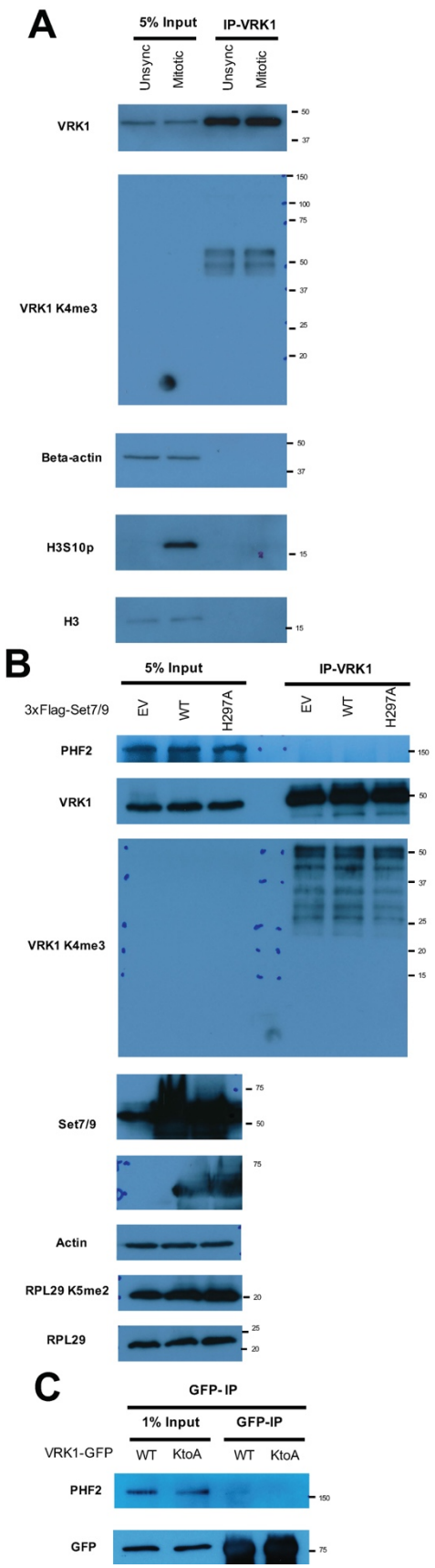


Figure 23. Mapping of VRK1 K4me3 under various circumstances. (A) 293T cells were treated with Nocodazole to induce mitotic arrest. VRK1 was immunoprecipitated from unsynchronized and mitotic cells and subjected to western blot analysis against VRK1 (Sigma), VRK1 K4me3, beta-actin, Histone H3S10 phosphorylation and H3. H3S10p served as a positive control for mitosis. (B) 293T cells were transiently transfected with empty vector (EV), 3xFlag-tagged WT and enzymatic-dead mutant H297A Set7/9. VRK1 was immunoprecipitated from three cell lines and subjected to western blot analysis against PHF2, VRK1 (Sigma), VRK1 K4me3, beta-actin, Set7/9, Flag, RPL29K5me2 and RPL29. (C) 293T cells were transiently transfected with WT and K4A mutant VRK1-GFP. VRK1-GFP was immunoprecipitated from three cell lines and subjected to western blot analysis against PHF2 and GFP.

CHAPTER 6: DISCUSSIONS AND FUTURE PLANS

Part of this chapter is based upon: J. Chen, , C. Sagum, and M.T. Bedford, (2020) Protein domain microarrays as a platform to decipher signaling pathways and the histone code. Methods, Volume 184: pp. 4-12.

ISSN 1046-2023

Used with permission of Elsevier

Part of this chapter is based upon: J. Chen, C. Sagum, J. Zhou, J. Horton J, X. Cheng[3] and M.T. Bedford, Histone H3 N-terminal mimicry drives a novel network of methyl-effector interactions. Under revision.

In **Chapter 3** of this dissertation, I discussed the history, design and application of protein domain microarray as a platform to decipher novel molecular interactions. In **Chapters 4** and **5**, I described application of two protein domain microarrays, ubiquitin-binding domain and methyl reader domain microarrays, to identify novel readers for histone H2Aub1 and H2Bub1, and H3TM K4me3, respectively.

Based on **Chapters 4** and **5**, protein domain microarray is a high-throughput screening platform that can probe various peptide probes harboring different PTMs for novel interactions with varying effector modules. At the same time, there are two main limitations of protein domain microarrays compared to the tandem affinity purification-mass spectrometry (TAP-MS) approach.

The first limitation is that protein domain microarrays require prior knowledge of reader domains, limiting novel interactions within the domains arrayed on the microchips and missing potential readers that haven't been previously characterized. For example, RSF1, an H2Aub1-specific reader identified using a TAP-MS approach, contains a previously uncharacterized ubiquitin-binding motif (Zhang et al., 2017b).

The second limitation is that most probes used with protein domain arrays are peptide probes that are prone to false-positive interactions. As both the peptide probes and protein domains on the microarrays are short fragments of full-length proteins, the identified novel interactions need to be carefully validated using *in vitro* and *in vivo* assays. For instance, in **Chapter 4**, HRS/HGS UIM domain was first identified as an H2Aub1-specific reader but didn't interact with the full-length H2Aub1 from the histone extracts, thus excluded from the following studies. In **Chapter 5**, I discovered several novel interactions between K4me3 H3TM peptides and

PHD fingers. After validation using pull-down assays, VRK1 K4me3-PHF2 PHD interaction was the most robust one, thus chosen for subsequent studies.

In addition, interactions that are not relatively strong (weaker than 30-50 μM KD) are not detected on the array. These weaker interactions can often be identified by the pull-down approaches. For example, our Bromo-domain microarray has difficulty detecting specific interactions that are readily identified by the pull-down approaches. Finally, most of the protein domains we have arrayed are generated in isolation, and not linked to any other domain that may occur in close linear proximity or even in the tight association through a protein complex. As a consequence, we lose the ability to identify multivalent PTM engagement. This is an emerging theme in the chromatin field, which unfortunately cannot be addressed by our arrays that generally do not contain more than one domain per spot.

There are a large number of different types of protein domains, and while a number of these domain types have been developed and used on array platforms, there are many more domain families that can be isolated, expressed as recombinant proteins and arrayed onto slides for detailed comparative analysis. I will highlight a few new directions below.

Develop DNA and RNA-binding protein domain arrays: A huge class of untapped protein domains belongs to domain families that bind nucleic acids, either DNA or RNA. In the former case, these primarily involve domains found in transcription factors (Garvie and Wolberger, 2001, Siggers and Gordan, 2014), and in the latter case, these are domains involved in splicing regulation and RNA shuttling (Lunde et al., 2007, Castello et al., 2016). There is also an emerging class of domains that bind both RNA and DNA (Hudson and Ortlund, 2014). These types of nucleic acid binding domain arrays hold the promise of not only identifying sequence-specific recognition protein modules, but also discovering selective readers of

modified DNA and RNA, and proteins that could bind double stranded RNA, single stranded DNA and RNA/DNA hybrids.

Develop Phospholipid-binding protein domain arrays: Phospholipids and their reading proteins are essential components in signal transduction, such as protein translocation and lipid modification, upon stimulation of cell surface receptors. Major phospholipids in mammalian membranes are phosphotidylserine, phosphatidic acid and phosphatidylinositol. Phosphatidylinositol can be phosphorylated at the 3-, 4- and/or 5- positions to generate phosphoinositides. The composition of phospholipids depends on cell type and is tightly regulated. There are more than ten phospholipid-binding domains (Lemmon, 2008). Like protein-protein interactions, multiple domains in the same protein can cooperate with each other to facilitate membrane bindings. Therefore, it is important to determine the binding affinity of a phospholipid-domain pair and avidity effects of multivalent multi-domain interactions. By arraying single and tandem phospholipid binding domains, phospholipid-binding domain arrays have the potential to dissect this sophisticated combinatorial “code”.

Develop ADP-ribose binding domain arrays: Poly(ADP-ribosyl)ation (PARylation) is an important PTM involved in chromatin dynamics, genome stability maintenance, transcription, cell metabolism and development (Teloni and Altmeyer, 2016). There are four major PAR-binding modules: PAR binding motif (PBM), Poly(ADP-Ribose)-binding Zinc Finger (PBZ), WWE domain, and macrodomain (Barkauskaite et al., 2013). ADP-ribose binding domain arrays will be promising in the identification of novel PARylation-dependent protein-protein interactions. The difficulty here will be the development of biotinylated PAR probes to facilitate the screening.

In **chapter 4** of this dissertation, I identified the PLAA PFU domain as a potential H2Bub1-specific effector module using a ubiquitin-binding domain array. However, the PLAA-H2Bub1 interaction may require additional factors for chromatin binding because the PFU domain specifically binds H2Bub1 from histone extracts but this does not translate into binding to nucleosomal recombinant H2Bub1. PTMs on the ubiquitin moiety of H2Bub1 could potentially contribute to the differences between endogenous H2Bub1 protein and recombinant H2Bub1 nucleosomes.

Endogenous H2Bub1 might harbor additional PTMs on the ubiquitin moiety, such as acetylation and phosphorylation, which regulate ubiquitin's conformation and binding characteristics (Ohtake et al., 2015, Kondapalli et al., 2012, Kane et al., 2014, Wauer et al., 2015, Aguirre et al., 2017, Swaney et al., 2015, Dong et al., 2017, Walser et al., 2020). Acetylation of ubiquitin at K6 (K6ac) inhibits elongation of the K11-, K48-, and K63-linked polyubiquitin chains (Ohtake et al., 2015). Gel regions covering monoubiquitinated histones were analyzed by LC-MS, identifying Histone H2A and H2B are substrates for acetyl-ubiquitinations, K6ac, and K48ac (Ohtake et al., 2015). Expression of K6Q (K6ac mimetic) ubiquitin increased H2Bub1 levels while that of K6R (non-acetylable) ubiquitin decreased H2Bub1 levels. Though there is no direct link between ubiquitin acetylation and H2Bub1 stability, the data suggests that H2Bub1 and H2Aub1 are subject to an additional layer of regulation by ubiquitin acetylation, especially in the context of active transcription, which is marked by H2Bub1 and high chromatin acetylation activity.

Similarly, ubiquitin phosphorylations at S65 (S65p Ub) and T12 (T12p Ub) modulate protein-protein interactions and downstream functions. Mediated by PINK1, S65p Ub functions as an allosteric activator for Parkin (Kondapalli et al., 2012, Kane et al., 2014, Wauer et al., 2015, Aguirre et al., 2017). The latter is an E3 ligase that orchestrates the turnover of damaged

mitochondria through mitophagy. Additionally, S65 phosphorylation allows ubiquitin to adopt two conformations in solution, a retracted state and a relaxed state (Dong et al., 2017, Swaney et al., 2015). The balance of these two states is pH-dependent, suggesting that S65 Ub might function in a location-specific manner, as pHs vary between separate cellular compartments (Casey et al., 2010).

A recent study found that T12p Ub on histone H2AK15ub is a novel histone mark that regulates DNA damage response (Walser et al., 2020). 53BP1 is an H2AK15ub-specific effector upon DNA damage that recruits other protein complexes to the DNA damage sites (Fradet-Turcotte et al., 2013). T12p Ub on H2AK12ub blocks 53BP1's chromatin association by preventing its binding to H2AK15ub. In addition, it is possible that the T12p Ub may serve as a docking site for an effector molecule that harbors a BRCT or FHA domain (two domain types that read pT/pS motifs). Indeed, preliminary data from our lab has identified one such reader.

All the above findings suggest that histone ubiquitination marks are subject to further modifications that orchestrate the conformation and interaction with effector modules. The next step is to characterize the H2Bub1 histone pulled down using the PLAA PFU domain and look for novel modifications on the ubiquitin moiety.

Accumulation of a 15 kDa Ub-like protein in PLAA knockout cells

I observed the accumulation of a 15 kDa band that can be recognized by the anti-Ubiquitin antibody in PLAA KO cells (Figure 10). One way to identify this protein is to analyze the gel bands at the 15 kDa regions by LC-MS, looking for highly abundant proteins in PLAA KO cells but not WT cells.

After ruling out the protein to be K48- and K63-linked di-ubiquitin, RPS27A, and ISG15, there is one last possible candidate, K29-linked di-ubiquitin. However, there are no commercially available antibodies against K29-Ub. Yu et al. recently described a synthetic antibody fragment that recognizes K29-Ub (Yu et al., 2020). Immunofluorescence showed K29-ub co-localizes with VCP and the 20S proteasome. More importantly, K29-ub is enriched around the midbody in telophase, the patterns of which are similar to that of PLAA observed in my current study (data not shown).

Given that Doa1 is a K29-Ub binder, PLAA might also bind K29-Ub chains and process them into mono-ubiquitin, which would explain the accumulation of the 15 kDa upon PLAA depletion.

This hypothesis could be tested either by LC-MS as described above or overexpression of K29R and WT ubiquitin in PLAA KO cells. Alternatively, we could reach out to Yu et al. to request their K29-Ub specific antibody fragment.

PLAA's recruitment to the DNA damage sites.

Together with 53BP1 and γH2AX, PLAA is localized to the laser irradiation-induced DNA damage sites (**Figure 12**), suggesting PLAA's function in DDR. Notably, H2Bub1 depletion was observed at DNA damage sites. The current study could not determine whether PLAA recruitment and H2Bub1 removal are related to each other.

As shown in this study and previously published work, PLAA interacts with VCP (Decottignies et al., 2004, Ghislain et al., 1996, Ogiso et al., 2004, Mullally et al., 2006). VCP is also involved in DNA damage response. VCP and its two cofactors, UFD1 and NPL4, mediate the extraction of ubiquitinated L3MBTL1 from H4K20me2 at DSBs, unmasking this mark for efficient 53BP1 recruitment (Acs et al., 2011).

The PLAA-VCP complex may function in a similar way by disengaging ubiquitinated proteins from DSBs. Considering that we observe the depletion of H2Bub1 at DSBs (**Figure 12B**), H2Bub1 is potentially one of the ubiquitinated substrates removed by the PLAA-VCP complex. More work is needed to test this hypothesis.

I propose three experiments to test these hypotheses: 1) Examine whether the recruitment of 53BP1 recruitment and depletion of H2Bub1 at DNA damage sites are defective in PLAA KO cells. 2) Rescue PLAA KO cells with PLAA truncations to map which domain(s) that is(are) required for PLAA enrichment at DSBs. 3) Identify the PLAA-interacting protein complex after DNA damage by LC-MS.

In summary, the study in **Chapter 4** identified PLAA as a potential H2Bub1-specific reader. I discovered the recruitment of PLAA recruited to the DNA damage sites and ubiquitin-related phenotypes in PLAA knockout cells, including reduced H2Aub1 and H2Bub1 and accumulation of a 15 kDa ubiquitin-like protein. Though additional work is needed, PLAA is possibly the first H2Bub1-specific effector protein.

In **Chapter 5** of this dissertation, I proposed the concept of histone H3 N-terminal mimicry and provided *in vitro* biochemical evidence that H3TM N-terminal K4me3 serves as a molecular switch that engage different protein interaction networks. These findings may dramatically expand the scope of functions that have been bestowed on methyl-effector proteins and the NuRD complex beyond the original histone code hypothesis by providing evidence that PHD fingers (and the Tudor domain of SPIN1) recognize not only histone proteins, but also non-histone proteins.

For these findings to be relevant, they of course need to also occur in cells and *in vivo*. However, we did not find direct evidence for the existence of the H3TM-K4me3 methylation state *in vivo*. This is not through lack of trying. First, using MS (**Table 6**), we were unable to map the N-terminal PTMs for all six tested H3TM candidates, because, like histone H3, the H3TMs are enriched for lysine or arginine residues in the vicinity of the K4 residue, generating tryptic peptides that are often very short and incompatible with MS analysis. Direct evidence for the occurrence of H3TM K4me3 will likely come from middle-/top-down mass spectrometry studies (Sidoli and Garcia, 2017). Second, we developed methyl-specific antibodies to K4me3 sites on VRK1, BCL11B, TSHZ1 and MEF2C. While these antibodies selectively recognize the K4me3 peptides, and not the K4me0 peptides, in dot-blot assays, they were unable to detect these three endogenous H3TMs. This could be due to a number of reasons: 1) the antibodies are just bad; 2) the H4me3 mark occurs at very low levels, which makes it difficult for the antibodies to detect; and 3) the H4me3 mark on four H3TMs may be cell-type specific, tissue specific, cell-cycle specific, or development specific, and we have missed them.

VRK1 is a nuclear kinase that regulates chromatin compaction in the G2/M phase (Valbuena et al., 2008, Jeong et al., 2013, Kang et al., 2007). I investigated whether the VRK1-K4me3, and the subsequent interaction with PHF2, could be related to VRK1's function in cell cycle regulation. However, there were no K4me3 signals or PHF2 interaction detected during mitotic arrest (**Figure 23**). Proteins often have functions that are yet to be discovered, limiting our understanding of the functional significance of protein-protein interactions beyond their classical roles. For instance, DNA ligase 1 (LIG1) K126me2/me3 has recently been identified as a novel ligand for the tandem Tudor domain of ubiquitin-like with PHD and RING finger domains 1 (UHRF1) using the same methyl reader domain microarray in **Chapter 5** (Vaughan et al., 2020). LIG1 K126me2/me3 and H3K9me2/me3 compete for UHRF1 interactions. However, UHRF1's interactions with H3K9 and LIG1 didn't affect UHRF1's function in

maintaining cancer cell DNA methylation levels, the most well-characterized function of UHRF1. UHRF1's functions as a non-histone methyl-lysine reader remain to be uncovered.

Although the cellular context of K4me3 remains unknown, the N-termini of H3TMs are disordered (https://iupred2a.elte.hu/) as is the H3 N-terminus, implying that the N-terminal regions need to be stabilized by PTMs or by interacting with binding partners (Bah and Forman-Kay, 2016). Indeed, H3TMs with unmodified N-terminal tails mimic unmodified histone H3. Several of the 48 H3TM candidates interact with the repressive NuRD complex, in their unmethylated state (Lauberth and Rauchman, 2006, Hong et al., 2005, Svensson et al., 2000, Dubuissez et al., 2016b, Moody et al., 2018, Cismasiu et al., 2005, Xu et al., 2013). Interestingly, the NuRD complex also recognizes the unmodified histone H3 N-terminus, but this interaction is disrupted by K4 trimethylation (Allen et al., 2013, Zegerman et al., 2002). K4me3 prohibits the chromatin-association of the repressive NuRD complex regulating local chromatin status to facilitate transcriptional activation. Among the 48 candidate proteins, Sall1/2/3/4, Fog1/2, and BCL11A/B are known to interact with the NuRD complex via their N-termini, repressing target gene expressions (Lauberth and Rauchman, 2006, Hong et al., 2005, Svensson et al., 2000, Dubuissez et al., 2016b, Moody et al., 2018, Cismasiu et al., 2005, Xu et al., 2013). However, the mechanism by which these interactions are turned off remains unknown. Among the seven candidates that were used to probe the protein domain microarray (**Figure 3**), VRK1 K4me1, MEF2C K4ac, TSHZ1 K4ac, and BCL11B S2ph have been identified in various mass spectrometry studies (Wu et al., 2015, Angelelli et al., 2008, Dubuissez et al., 2016a, Schwer et al., 2009). Notably, BCL11B serine 2 phosphorylation has been shown to negatively regulate BCL11B's interaction with the NuRD complex (Dubuissez et al., 2016b). Given that the N-terminal tail of histone H3 is heavily modified by various PTMs, such as phosphorylation, methylation, and acetylation, it is likely that the N-termini of these

H3TM proteins are also subject to PTMs that positively or negatively regulate their ability to interact with recognition modules.

We have proposed a model that H3TMs bind the repressive NuRD complex in their unmethylated (K4me0) state, which is already well established, and in their K4me3 state these same H3TMs directly recruit effectors (protein that harbor PHD figures and Tudor domains) that have transcriptional activator activity (**Figure 24**). In cases where the binding affinity of the effector is greater for the H3TM than that for H3K4me3, these effectors may already reside on active chromatin and then be transferred to the H3TM. If the binding affinities are reversed (higher for H3K4me3 than H3TM-Kme3) then the H3TM may "deliver" the effector to chromatin.

A search for proteins that harbor N-termini with sequence similarity to histone H3 identified striking enrichment (5-fold) of proteins with chromatin binding functions (**Figure 14**). Many of these H3TM proteins have already be shown to interact with the NuRD transcriptional repressor complex in their unmethylated states. We tested 7 of these methylated H3TMs for their ability to interact with known histone H3K4me3 binders, and they all do, to varying degrees. We thus hypothesize that the 48 H3TMs that we identified will be a major class of interacting proteins for what are traditionally thought of as H3K4me effector proteins.

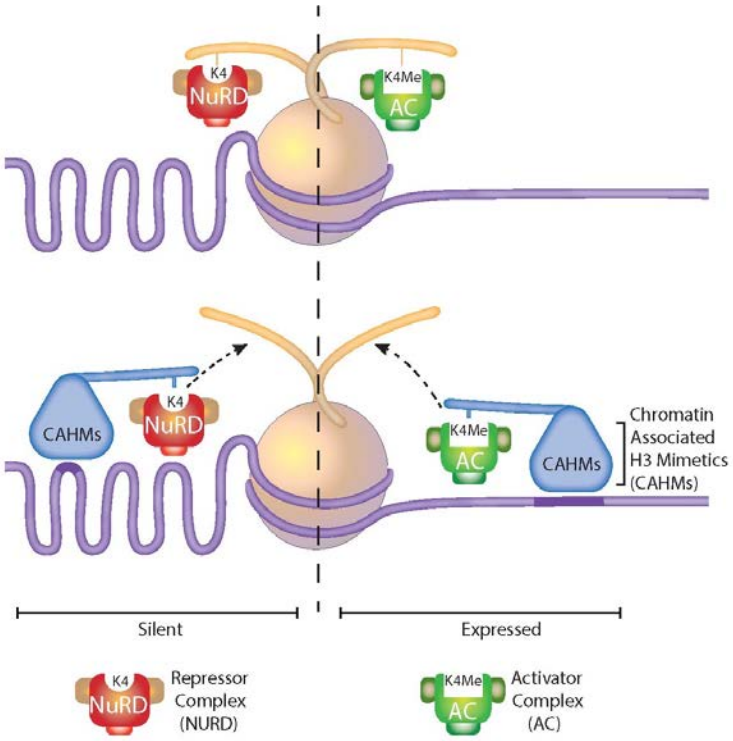


Figure 24. Regulation of H3TMs' interaction networks by K4me3. Proteins containing PHD or Tudor domains often recognize the Histone H3K4me3, an active transcription mark, and recruit associating activator protein complex (AC) to the chromatin. Repressor complex, such as NuRD, interacts with chromatin by recognizing the unmodified Histone H3 tail. When unmodified at K4 sites, H3TMs interact with the NuRD complex. K4me3 on H3TMs disengages their interactions with the NuRD complex and turns on their interactions with AC to execute H3TMs' nuclear functions.

BIBLIOGRAPHY

AASLAND, R., GIBSON, T. J. & STEWART, A. F. 1995. The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends in Biochemical Sciences,* 20**,** 56-59.

ACS, K., LUIJSTERBURG, M. S., ACKERMANN, L., SALOMONS, F. A., HOPPE, T. & DANTUMA, N. P. 2011. The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks. *Nature Structural & Molecular Biology,* 18**,** 1345-1350.

AGUIRRE, J. D., DUNKERLEY, K. M., MERCIER, P. & SHAW, G. S. 2017. Structure of phosphorylated UBL domain and insights into PINK1-orchestrated parkin activation. *Proceedings of the National Academy of Sciences,* 114**,** 298-303.

ALLEN, H. F., WADE, P. A. & KUTATELADZE, T. G. 2013. The NuRD architecture. *Cellular and Molecular Life Sciences,* 70**,** 3513-3524.

ANGELELLI, C., MAGLI, A., FERRARI, D., GANASSI, M., MATAFORA, V., PARISE, F., RAZZINI, G., BACHI, A., FERRARI, S. & MOLINARI, S. 2008. Differentiation-dependent lysine 4 acetylation enhances MEF2C binding to DNA in skeletal muscle cells. *Nucleic acids research,* 36**,** 915-928.

ATANASSOV, B. S., MOHAN, R. D., LAN, X., KUANG, X., LU, Y., LIN, K., MCIVOR, E., LI, W., ZHANG, Y. & FLORENS, L. 2016. ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. *Molecular cell,* 62**,** 558-571.

BADEAUX, A. I. & SHI, Y. 2013. Emerging roles for chromatin as a signal integration and storage platform. *Nat Rev Mol Cell Biol,* 14**,** 211-24.

BAH, A. & FORMAN-KAY, J. D. 2016. Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry,* 291**,** 6696-6705.

BANNISTER, A. J., ZEGERMAN, P., PARTRIDGE, J. F., MISKA, E. A., THOMAS, J. O., ALLSHIRE, R. C. & KOUZARIDES, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* 410**,** 120-124.

BARKAUSKAITE, E., JANKEVICIUS, G., LADURNER, A. G., AHEL, I. & TIMINSZKY, G. 2013. The recognition and removal of cellular poly(ADP-ribose) signals. *FEBS J,* 280**,** 3491-507.

BELLARE, P., SMALL, E. C., HUANG, X., WOHLSCHLEGEL, J. A., STALEY, J. P. & SONTHEIMER, E. J. 2008. A role for ubiquitin in the spliceosome assembly pathway. *Nat Struct Mol Biol,* 15**,** 444-51.

BERNSTEIN, B. E., KAMAL, M., LINDBLAD-TOH, K., BEKIRANOV, S., BAILEY, D. K., HUEBERT, D. J., MCMAHON, S., KARLSSON, E. K., KULBOKAS III, E. J. & GINGERAS, T. R. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell,* 120**,** 169-181.

BERTOLAET, B. L., CLARKE, D. J., WOLFF, M., WATSON, M. H., HENZE, M., DIVITA, G. & REED, S. I. 2001. UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nat Struct Biol,* 8**,** 417-22.

BHATTACHARYYA, R. P., REMÉNYI, A., GOOD, M. C., BASHOR, C. J., FALICK, A. M. & LIM, W. A. 2006. The Ste5 Scaffold Allosterically Modulates Signaling Output of the Yeast Mating Pathway. *Science,* 311**,** 822-826.

BIENKO, M., GREEN, C. M., CROSETTO, N., RUDOLF, F., ZAPART, G., COULL, B., KANNOUCHE, P., WIDER, G., PETER, M., LEHMANN, A. R., HOFMANN, K. & DIKIC, I. 2005. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science,* 310**,** 1821-4.

BIENZ, M. 2006. The PHD finger, a nuclear protein-interaction domain. *Trends in biochemical sciences,* 31**,** 35-40.

BLOMSTROM, D. C., FAHEY, D., KUTNY, R., KORANT, B. D. & KNIGHT JR, E. 1986. Molecular characterization of the interferon-induced 15-kDa protein. Molecular cloning and nucleotide and amino acid sequence. *Journal of Biological Chemistry,* 261**,** 8811-8816.

BOMALASKI, J. S., FALLON, M., TURNER, R. A., CROOKE, S. T., MEUNIER, P. C. & CLARK, M. A. 1990. Identification and isolation of a phospholipase A2 activating protein in human rheumatoid arthritis synovial fluid: induction of eicosanoid synthesis and an inflammatory response in joints injected in vivo. *The Journal of laboratory and clinical medicine,* 116**,** 814-825.

BRADSHAW, R. A., BRICKEY, W. W. & WALKER, K. W. 1998. N-terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends in biochemical sciences,* 23**,** 263-267.

BRANDT, W., STRICKLAND, W., MORGAN, M. & VON HOLT, C. 1974. Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant. *FEBS letters,* 40**,** 167-172.

BUCHBERGER, A., SCHINDELIN, H. & HÄNZELMANN, P. 2015. Control of p97 function by cofactor binding. *FEBS letters,* 589**,** 2578-2589.

BUETOW, L. & HUANG, D. T. 2016. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nature reviews Molecular cell biology,* 17**,** 626.

BUSSOW, K., CAHILL, D., NIETFELD, W., BANCROFT, D., SCHERZINGER, E., LEHRACH, H. & WALTER, G. 1998. A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res,* 26**,** 5007-5008.

CASEY, J. R., GRINSTEIN, S. & ORLOWSKI, J. 2010. Sensors and regulators of intracellular pH. *Nature reviews Molecular cell biology,* 11**,** 50-61.

CASTELLO, A., FISCHER, B., FRESE, C. K., HOROS, R., ALLEAUME, A. M., FOEHR, S., CURK, T., KRIJGSVELD, J. & HENTZE, M. W. 2016. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell,* 63**,** 696-710.

CHAKRAVARTY, S., ZENG, L. & ZHOU, M.-M. 2009. Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure,* 17**,** 670-679.

CHAN, D. C., BEDFORD, M. T. & LEDER, P. 1996. Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. *Embo J,* 15**,** 1045-54.

CHANG, I. F. 2006. Mass spectrometry‐based proteomic analysis of the epitope‐tag affinity purified protein complexes in eukaryotes. *Proteomics,* 6**,** 6158-6166.

CHEN, J., SAGUM, C. & BEDFORD, M. T. 2019. Protein domain microarrays as a platform to decipher signaling pathways and the histone code. *Methods*.

CHEN, J., SAGUM, C. & BEDFORD, M. T. 2020. Protein domain microarrays as a platform to decipher signaling pathways and the histone code. *Methods,* 184**,** 4-12.

CHEN, S., YANG, Z., WILKINSON, ALEX W., DESHPANDE, ANIRUDDHA J., SIDOLI, S., KRAJEWSKI, K., STRAHL, BRIAN D., GARCIA, BENJAMIN A., ARMSTRONG, SCOTT A., PATEL, DINSHAW J. & GOZANI, O. 2015. The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. *Molecular Cell,* 60**,** 319-327.

CHENG, D., CÔTÉ, J., SHAABAN, S. & BEDFORD, M. T. 2007. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Molecular cell,* 25**,** 71-83.

CHIGNOLA, F., GAETANI, M., REBANE, A., ORG, T., MOLLICA, L., ZUCCHELLI, C., SPITALERI, A., MANNELLA, V., PETERSON, P. & MUSCO, G. 2009. The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic acids research,* 37**,** 2951-2961.

CICCHETTI, P., MAYER, B. J., THIEL, G. & BALTIMORE, D. 1992. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science,* 257**,** 803-6.

CISMASIU, V. B., ADAMO, K., GECEWICZ, J., DUQUE, J., LIN, Q. & AVRAM, D. 2005. BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene,* 24**,** 6753-6764.

CLARK, M. A., CHEN, M., CROOKE, S. T. & BOMALASKI, J. S. 1988. Tumour necrosis factor (cachectin) induces phospholipase A2 activity and synthesis of a phospholipase A2-activating protein in endothelial cells. *Biochemical Journal,* 250**,** 125-132.

CLARK, M. A., CONWAY, T. M., SHORR, R. & CROOKE, S. T. 1987. Identification and isolation of a mammalian protein which is antigenically and functionally related to the phospholipase A2 stimulatory peptide melittin. *Journal of Biological Chemistry,* 262**,** 4402-4406.

CLARK, M. A., OZGÜR, L., CONWAY, T. M., DISPOTO, J., CROOKE, S. T. & BOMALASKI, J. S. 1991. Cloning of a phospholipase A2-activating protein. *Proceedings of the National Academy of Sciences,* 88**,** 5418-5422.

COLLINS, R. E., NORTHROP, J. P., HORTON, J. R., LEE, D. Y., ZHANG, X., STALLCUP, M. R. & CHENG, X. 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol,* 15**,** 245-50.

COOPER, S., DIENSTBIER, M., HASSAN, R., SCHERMELLEH, L., SHARIF, J., BLACKLEDGE, N. P., DE MARCO, V., ELDERKIN, S., KOSEKI, H. & KLOSE, R. 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell reports,* 7**,** 1456-1470.

COOPER, S., GRIJZENHOUT, A., UNDERWOOD, E., ANCELIN, K., ZHANG, T., NESTEROVA, T. B., ANIL-KIRMIZITAS, B., BASSETT, A., KOOISTRA, S. M. & AGGER, K. 2016a. Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature communications,* 7**,** 1-8.

COOPER, S., GRIJZENHOUT, A., UNDERWOOD, E., ANCELIN, K., ZHANG, T., NESTEROVA, T. B., ANIL-KIRMIZITAS, B., BASSETT, A., KOOISTRA, S. M., AGGER, K., HELIN, K., HEARD, E. & BROCKDORFF, N. 2016b. Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature Communications,* 7**,** 13661.

CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-563.

DAI, R., FREJTAG, W., HE, B., ZHANG, Y. & MIVECHI, N. F. 2000. c-Jun NH2-terminal kinase targeting and phosphorylation of heat shock factor-1 suppress its transcriptional activity. *Journal of Biological Chemistry,* 275**,** 18210-18218.

DECOTTIGNIES, A., EVAIN, A. & GHISLAIN, M. 2004. Binding of Cdc48p to a ubiquitin-related UBX domain from novel yeast proteins involved in intracellular proteolysis and sporulation. *Yeast,* 21**,** 127-139.

DERIBE, Y. L., PAWSON, T. & DIKIC, I. 2010a. Post-translational modifications in signal integration. *Nat Struct Mol Biol,* 17**,** 666-72.

DERIBE, Y. L., PAWSON, T. & DIKIC, I. 2010b. Post-translational modifications in signal integration. *Nature structural & molecular biology,* 17**,** 666.

DEVERAUX, Q., USTRELL, V., PICKART, C. & RECHSTEINER, M. 1994. A 26 S protease subunit that binds ubiquitin conjugates. *J Biol Chem,* 269**,** 7059-61.

DEVINE, T., SEARS, R. C. & DAI, M.-S. 2018. The ubiquitin-specific protease USP36 is a conserved histone H2B deubiquitinase. *Biochemical and biophysical research communications,* 495**,** 2363-2368.

DIKIC, I., WAKATSUKI, S. & WALTERS, K. J. 2009. Ubiquitin-binding domains - from structures to functions. *Nat Rev Mol Cell Biol,* 10**,** 659-71.

DONG, X., GONG, Z., LU, Y.-B., LIU, K., QIN, L.-Y., RAN, M.-L., ZHANG, C.-L., LIU, Z., ZHANG, W.-P. & TANG, C. 2017. Ubiquitin S65 phosphorylation engenders a pH-sensitive conformational switch. *Proceedings of the National Academy of Sciences,* 114**,** 6770-6775.

DOU, Y., MILNE, T. A., RUTHENBURG, A. J., LEE, S., LEE, J. W., VERDINE, G. L., ALLIS, C. D. & ROEDER, R. G. 2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature Structural & Molecular Biology,* 13**,** 713-719.

DUBUISSEZ, M., LOISON, I., PAGET, S., VORNG, H., AIT-YAHIA, S., ROHR, O., TSICOPOULOS, A. & LEPRINCE, D. 2016a. Protein kinase C-mediated phosphorylation of BCL11B at serine 2 negatively regulates its interaction with NuRD complexes during CD4+ T-cell activation. *Molecular and cellular biology,* 36**,** 1881-1898.

DUBUISSEZ, M., LOISON, I., PAGET, S., VORNG, H., AIT-YAHIA, S., ROHR, O., TSICOPOULOS, A. & LEPRINCE, D. 2016b. Protein Kinase C-Mediated Phosphorylation of BCL11B at Serine 2 Negatively Regulates Its Interaction

with NuRD Complexes during CD4⁺ T-Cell Activation. *Molecular and Cellular Biology,* 36**,** 1881-1898.

ESPEJO, A. & BEDFORD, M. T. 2004a. Protein-domain microarrays. *Protein Arrays.* Springer.

ESPEJO, A. & BEDFORD, M. T. 2004b. Protein-domain microarrays. *Methods Mol Biol,* 264**,** 173-81.

ESPEJO, A., COTE, J., BEDNAREK, A., RICHARD, S. & BEDFORD, M. T. 2002. A protein-domain microarray identifies novel protein-protein interactions. *Biochem J,* 367**,** 697-702.

FALIK ZACCAI, T. C., SAVITZKI, D., ZIVONY-ELBOUM, Y., VILBOUX, T., FITTS, E. C., SHOVAL, Y., KALFON, L., SAMRA, N., KEREN, Z., GROSS, B., CHASNYK, N., STRAUSSBERG, R., MULLIKIN, J. C., TEER, J. K., GEIGER, D., KORNITZER, D., BITTERMAN-DEUTSCH, O., SAMSON, A. O., WAKAMIYA, M., PETERSON, J. W., KIRTLEY, M. L., PINCHUK, I. V., BAZE, W. B., GAHL, W. A., KLETA, R., ANIKSTER, Y. & CHOPRA, A. K. 2016. Phospholipase A2-activating protein is associated with a novel form of leukoencephalopathy. *Brain,* 140**,** 370-386.

FIEDLER, M., SÁNCHEZ-BARRENA, M. J., NEKRASOV, M., MIESZCZANEK, J., RYBIN, V., MÜLLER, J., EVANS, P. & BIENZ, M. 2008. Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Molecular cell,* 30**,** 507-518.

FIERZ, B., CHATTERJEE, C., MCGINTY, R. K., BAR-DAGAN, M., RALEIGH, D. P. & MUIR, T. W. 2011. Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nature chemical biology,* 7**,** 113-119.

FLANAGAN, J. F., MI, L.-Z., CHRUSZCZ, M., CYMBOROWSKI, M., CLINES, K. L., KIM, Y., MINOR, W., RASTINEJAD, F. & KHORASANIZADEH, S. 2005. Double

chromodomains cooperate to recognize the methylated histone H3 tail. *Nature,* 438**,** 1181-1185.

FLEMING, Y., ARMSTRONG, C. G., MORRICE, N., PATERSON, A., GOEDERT, M. & COHEN, P. 2000. Synergistic activation of stress-activated protein kinase 1/c-Jun N-terminal kinase (SAPK1/JNK) isoforms by mitogen-activated protein kinase kinase 4 (MKK4) and MKK7. *Biochemical Journal,* 352**,** 145-154.

FRADET-TURCOTTE, A., CANNY, M. D., ESCRIBANO-DÍAZ, C., ORTHWEIN, A., LEUNG, C. C., HUANG, H., LANDRY, M.-C., KITEVSKI-LEBLANC, J., NOORDERMEER, S. M. & SICHERI, F. 2013. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature,* 499**,** 50-54.

FROTTIN, F., MARTINEZ, A., PEYNOT, P., MITRA, S., HOLZ, R. C., GIGLIONE, C. & MEINNEL, T. 2006. The proteomics of N-terminal methionine cleavage. *Molecular & Cellular Proteomics,* 5**,** 2336-2349.

FU, Q.-S., ZHOU, C.-J., GAO, H.-C., JIANG, Y.-J., ZHOU, Z.-R., HONG, J., YAO, W.-M., SONG, A.-X., LIN, D.-H. & HU, H.-Y. 2009a. Structural basis for ubiquitin recognition by a novel domain from human phospholipase A2-activating protein. *Journal of Biological Chemistry,* 284**,** 19043-19052.

FU, Q. S., ZHOU, C. J., GAO, H. C., JIANG, Y. J., ZHOU, Z. R., HONG, J., YAO, W. M., SONG, A. X., LIN, D. H. & HU, H. Y. 2009b. Structural basis for ubiquitin recognition by a novel domain from human phospholipase A2-activating protein. *J Biol Chem,* 284**,** 19043-52.

FUKUNAGA, R. & HUNTER, T. 1997. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *Embo J,* 16**,** 1921-33.

GARVIE, C. W. & WOLBERGER, C. 2001. Recognition of specific DNA sequences. *Mol Cell,* 8**,** 937-46.

GHISLAIN, M., DOHMEN, R. J., LEVY, F. & VARSHAVSKY, A. 1996. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in Saccharomyces cerevisiae. *The EMBO journal,* 15**,** 4884-4899.

GIGLIONE, C., BOULAROT, A. & MEINNEL, T. 2004. Protein N-terminal methionine excision. *Cellular and Molecular Life Sciences CMLS,* 61**,** 1455-1474.

GOLDKNOPF, I., TAYLOR, C. W., BAUM, R. M., YEOMAN, L. C., OLSON, M., PRESTAYKO, A. & BUSCH, H. 1975. Isolation and characterization of protein A24, a" histone-like" non-histone chromosomal protein. *Journal of Biological Chemistry,* 250**,** 7182-7187.

GOLDKNOPF, I. L. & BUSCH, H. 1977. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proceedings of the National Academy of Sciences,* 74**,** 864-868.

GOLDKNOPF, I. L., FRENCH, M. F., MUSSO, R. & BUSCH, H. 1977. Presence of protein A24 in rat liver nucleosomes. *Proc Natl Acad Sci U S A,* 74**,** 5492-5.

GONG, F., CHIU, L.-Y., COX, B., AYMARD, F., CLOUAIRE, T., LEUNG, J. W., CAMMARATA, M., PEREZ, M., AGARWAL, P. & BRODBELT, J. S. 2015. Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes & development,* 29**,** 197-211.

GRAY, F., CHO, H. J., SHUKLA, S., HE, S., HARRIS, A., BOYTSOV, B., JAREMKO, Ł., JAREMKO, M., DEMELER, B. & LAWLOR, E. R. 2016. BMI1 regulates PRC1 architecture and activity through homo-and hetero-oligomerization. *Nature communications,* 7**,** 1-12.

GU, Y., JONES, A. E., YANG, W., LIU, S., DAI, Q., LIU, Y., SWINDLE, C. S., ZHOU, D., ZHANG, Z. & RYAN, T. M. 2016. The histone H2A deubiquitinase Usp16 regulates hematopoiesis and hematopoietic stem cell function. *Proceedings of the National Academy of Sciences,* 113**,** E51-E60.

GUPTA, S., BARRETT, T., WHITMARSH, A. J., CAVANAGH, J., SLUSS, H. K., DERIJARD, B. & DAVIS, R. J. 1996. Selective interaction of JNK protein kinase isoforms with transcription factors. *The EMBO journal,* 15**,** 2760-2770.

GUTIERREZ, C. B., YU, C., NOVITSKY, E. J., HUSZAGH, A. S., RYCHNOVSKY, S. D. & HUANG, L. 2016. Developing an acidic residue reactive and sulfoxide-containing MS-cleavable homobifunctional cross-linker for probing protein–protein interactions. *Analytical chemistry,* 88**,** 8315-8322.

HALL, E. A., NAHORSKI, M. S., MURRAY, L. M., SHAHEEN, R., PERKINS, E., DISSANAYAKE, K. N., KRISTARYANTO, Y., JONES, R. A., VOGT, J. & RIVAGORDA, M. 2017. PLAA mutations cause a lethal infantile epileptic encephalopathy by disrupting ubiquitin-mediated endolysosomal degradation of synaptic proteins. *The American Journal of Human Genetics,* 100**,** 706-724.

HAMIDI, T., SINGH, A. K., VELAND, N., VEMULAPALLI, V., CHEN, J., HARDIKAR, S., BAO, J., FRY, C. J., YANG, V. & LEE, K. A. 2018. Identification of Rpl29 as a major substrate of the lysine methyltransferase Set7/9. *Journal of Biological Chemistry,* 293**,** 12770-12780.

HE, L., LIU, X., YANG, J., LI, W., LIU, S., LIU, X., YANG, Z., REN, J., WANG, Y. & SHAN, L. 2018. Imbalance of the reciprocally inhibitory loop between the ubiquitin-specific protease USP43 and EGFR/PI3K/AKT drives breast carcinogenesis. *Cell research,* 28**,** 934-951.

HE, M. & TAUSSIG, M. J. 2001. Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Res,* 29**,** E73-3.

HJERPE, R., AILLET, F., LOPITZ-OTSOA, F., LANG, V., ENGLAND, P. & RODRIGUEZ, M. S. 2009. Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO reports,* 10**,** 1250-1258.

HOCK, A. K., VIGNERON, A. M. & VOUSDEN, K. H. 2014. Ubiquitin-specific peptidase 42 (USP42) functions to deubiquitylate histones and regulate transcriptional activity. *Journal of Biological Chemistry,* 289**,** 34862-34870.

HONG, W., NAKAZAWA, M., CHEN, Y.-Y., KORI, R., VAKOC, C. R., RAKOWSKI, C. & BLOBEL, G. A. 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *The EMBO journal,* 24**,** 2367-2378.

HOOK, S. S., ORIAN, A., COWLEY, S. M. & EISENMAN, R. N. 2002. Histone deacetylase 6 binds polyubiquitin through its zinc finger (PAZ domain) and copurifies with deubiquitinating enzymes. *Proc Natl Acad Sci U S A,* 99**,** 13425-30.

HORTON, J. R., UPADHYAY, A. K., HASHIMOTO, H., ZHANG, X. & CHENG, X. 2011. Structural basis for human PHF2 Jumonji domain interaction with metal ions. *Journal of molecular biology,* 406**,** 1-8.

HUANG, Y., FANG, J., BEDFORD, M. T., ZHANG, Y. & XU, R.-M. 2006a. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science,* 312**,** 748-751.

HUANG, Y., FANG, J., BEDFORD, M. T., ZHANG, Y. & XU, R. M. 2006b. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science,* 312**,** 748-51.

HUDSON, W. H. & ORTLUND, E. A. 2014. The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol,* 15**,** 749-60.

HUNG, T., BINDA, O., CHAMPAGNE, K. S., KUO, A. J., JOHNSON, K., CHANG, H. Y., SIMON, M. D., KUTATELADZE, T. G. & GOZANI, O. 2009. ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Molecular cell,* 33**,** 248-256.

HUYEN, Y., ZGHEIB, O., DITULLIO, R. A., JR., GORGOULIS, V. G., ZACHARATOS, P., PETTY, T. J., SHESTON, E. A., MELLERT, H. S., STAVRIDI, E. S. & HALAZONETIS, T. D. 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature,* 432**,** 406-11.

JAIN, K., FRASER, C. S., MARUNDE, M. R., PARKER, M. M., SAGUM, C., BURG, J. M., HALL, N., POPOVA, I. K., RODRIGUEZ, K. L., VAIDYA, A., KRAJEWSKI, K., KEOGH, M. C., BEDFORD, M. T. & STRAHL, B. D. 2020. Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics Chromatin,* 13**,** 3.

JENUWEIN, T. & ALLIS, C. D. 2001a. Translating the histone code. *Science,* 293**,** 1074-1080.

JENUWEIN, T. & ALLIS, C. D. 2001b. Translating the histone code. *Science,* 293**,** 1074-80.

JEONG, J. S., JIANG, L., ALBINO, E., MARRERO, J., RHO, H. S., HU, J., HU, S., VERA, C., BAYRON-POUEYMIROY, D., RIVERA-PACHECO, Z. A., RAMOS, L., TORRES-CASTRO, C., QIAN, J., BONAVENTURA, J., BOEKE, J. D., YAP, W. Y., PINO, I., EICHINGER, D. J., ZHU, H. & BLACKSHAW, S. 2012. Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. *Mol Cell Proteomics,* 11**,** O111 016253.

JEONG, M.-W., KANG, T.-H., KIM, W., CHOI, Y. H. & KIM, K.-T. 2013. Mitogen-activated protein kinase phosphatase 2 regulates histone H3 phosphorylation via

interaction with vaccinia-related kinase 1. *Molecular biology of the cell,* 24**,** 373-384.

JOHNSON, E. S., BARTEL, B., SEUFERT, W. & VARSHAVSKY, A. 1992. Ubiquitin as a degradation signal. *The EMBO journal,* 11**,** 497-505.

JOHNSON, E. S., MA, P. C., OTA, I. M. & VARSHAVSKY, A. 1995. A proteolytic pathway that recognizes ubiquitin as a degradation signal. *Journal of biological chemistry,* 270**,** 17442-17456.

KAIS, Z., BARSKY, S. H., MATHSYARAJA, H., ZHA, A., RANSBURGH, D. J., HE, G., PILARSKI, R. T., SHAPIRO, C. L., HUANG, K. & PARVIN, J. D. 2011. KIAA0101 interacts with BRCA1 and regulates centrosome number. *Molecular Cancer Research,* 9**,** 1091-1099.

KAKAROUGKAS, A., ISMAIL, A., CHAMBERS, A. L., RIBALLO, E., HERBERT, A. D., KÜNZEL, J., LÖBRICH, M., JEGGO, P. A. & DOWNS, J. A. 2014. Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Molecular cell,* 55**,** 723-732.

KALB, R., LATWIEL, S., BAYMAZ, H. I., JANSEN, P. W., MÜLLER, C. W., VERMEULEN, M. & MÜLLER, J. 2014a. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature structural & molecular biology,* 21**,** 569-571.

KALB, R., LATWIEL, S., BAYMAZ, H. I., JANSEN, P. W. T. C., MÜLLER, C. W., VERMEULEN, M. & MÜLLER, J. 2014b. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature Structural & Molecular Biology,* 21**,** 569-571.

KALB, R., MALLERY, D. L., LARKIN, C., HUANG, J. T. & HIOM, K. 2014c. BRCA1 is a histone-H2A-specific ubiquitin ligase. *Cell reports,* 8**,** 999-1005.

KANE, L. A., LAZAROU, M., FOGEL, A. I., LI, Y., YAMANO, K., SARRAF, S. A., BANERJEE, S. & YOULE, R. J. 2014. PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *Journal of Cell Biology,* 205**,** 143-153.

KANG, T.-H., PARK, D.-Y., CHOI, Y. H., KIM, K.-J., YOON, H. S. & KIM, K.-T. 2007. Mitotic histone H3 phosphorylation by vaccinia-related kinase 1 in mammalian cells. *Molecular and cellular biology,* 27**,** 8533-8546.

KIM, J., DANIEL, J., ESPEJO, A., LAKE, A., KRISHNA, M., XIA, L., ZHANG, Y. & BEDFORD, M. T. 2006a. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO reports,* 7**,** 397-403.

KIM, J., DANIEL, J., ESPEJO, A., LAKE, A., KRISHNA, M., XIA, L., ZHANG, Y. & BEDFORD, M. T. 2006b. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep,* 7**,** 397-403.

KIM, J., GUERMAH, M., MCGINTY, R. K., LEE, J.-S., TANG, Z., MILNE, T. A., SHILATIFARD, A., MUIR, T. W. & ROEDER, R. G. 2009. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell,* 137**,** 459-471.

KIM, S.-A., ZHU, J., YENNAWAR, N., EEK, P. & TAN, S. 2020. Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. *Molecular Cell,* 78**,** 903-914.e4.

KOMANDER, D. & RAPE, M. 2012. The ubiquitin code. *Annual review of biochemistry,* 81**,** 203-229.

KONDAPALLI, C., KAZLAUSKAITE, A., ZHANG, N., WOODROOF, H. I., CAMPBELL, D. G., GOURLAY, R., BURCHELL, L., WALDEN, H., MACARTNEY, T. J., DEAK, M., KNEBEL, A., ALESSI, D. R. & MUQIT, M. M. K. 2012. PINK1 is activated by mitochondrial membrane potential depolarization and stimulates

Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biology,* 2**,** 120080.

KORANT, B., BLOMSTROM, D. C., JONAK, G. J. & KNIGHT JR, E. 1984. Interferon-induced proteins. Purification and characterization of a 15,000-dalton protein from human and bovine cells induced by interferon. *Journal of Biological Chemistry,* 259**,** 14835-14839.

KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

KUO, A. J., SONG, J., CHEUNG, P., ISHIBE-MURAKAMI, S., YAMAZOE, S., CHEN, J. K., PATEL, D. J. & GOZANI, O. 2012. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature,* 484**,** 115-9.

KUROKAWA, M., MITANI, K., YAMAGATA, T., TAKAHASHI, T., IZUTSU, K., OGAWA, S., MORIGUCHI, T., NISHIDA, E., YAZAKI, Y. & HIRAI, H. 2000. The evi-1 oncoprotein inhibits c-Jun N-terminal kinase and prevents stress-induced cell death. *The EMBO journal,* 19**,** 2958-2968.

LACHNER, M., O'CARROLL, D., REA, S., MECHTLER, K. & JENUWEIN, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature,* 410**,** 116-20.

LAN, F., COLLINS, R. E., DE CEGLI, R., ALPATOV, R., HORTON, J. R., SHI, X., GOZANI, O., CHENG, X. & SHI, Y. 2007. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature,* 448**,** 718-722.

LAN, X., ATANASSOV, B. S., LI, W., ZHANG, Y., FLORENS, L., MOHAN, R. D., GALARDY, P. J., WASHBURN, M. P., WORKMAN, J. L. & DENT, S. Y.

2016. USP44 is an integral component of N-CoR that contributes to gene repression by deubiquitinating histone H2B. *Cell reports,* 17**,** 2382-2393.

LAUBERTH, S. M. & RAUCHMAN, M. 2006. A Conserved 12-Amino Acid Motif in Sall1 Recruits the Nucleosome Remodeling and Deacetylase Corepressor Complex. *Journal of Biological Chemistry,* 281**,** 23922-23931.

LEE, J. & BEDFORD, M. T. 2002. PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Rep,* 3**,** 268-73.

LEMMON, M. A. 2008. Membrane recognition by phospholipid-binding domains. *Nat Rev Mol Cell Biol,* 9**,** 99-111.

LEVY, D., KUO, A. J., CHANG, Y., SCHAEFER, U., KITSON, C., CHEUNG, P., ESPEJO, A., ZEE, B. M., LIU, C. L., TANGSOMBATVISIT, S., TENNEN, R. I., KUO, A. Y., TANJING, S., CHEUNG, R., CHUA, K. F., UTZ, P. J., SHI, X., PRINJHA, R. K., LEE, K., GARCIA, B. A., BEDFORD, M. T., TARAKHOVSKY, A., CHENG, X. & GOZANI, O. 2011. Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. *Nat Immunol,* 12**,** 29-36.

LI, C., IRRAZABAL, T., SO, C. C., BERRU, M., DU, L., LAM, E., LING, A. K., GOMMERMAN, J. L., PAN-HAMMARSTRÖM, Q. & MARTIN, A. 2018. The H2B deubiquitinase Usp22 promotes antibody class switch recombination by facilitating non-homologous end joining. *Nature Communications,* 9**,** 1006.

LI, X. S., TROJER, P., MATSUMURA, T., TREISMAN, J. E. & TANESE, N. 2010. Mammalian SWI/SNF-A subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Molecular and cellular biology,* 30**,** 1673-1688.

LI, Z., CAO, R., WANG, M., MYERS, M. P., ZHANG, Y. & XU, R.-M. 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *Journal of Biological Chemistry,* 281**,** 20643-20649.

LIS, E. T. & ROMESBERG, F. E. 2006. Role of Doa1 in the Saccharomyces cerevisiae DNA damage response. *Molecular and cellular biology,* 26**,** 4122-4133.

LIU, F., RIJKERS, D. T., POST, H. & HECK, A. J. 2015. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature methods,* 12**,** 1179-1184.

LOBINGIER, B. T., HÜTTENHAIN, R., EICHEL, K., MILLER, K. B., TING, A. Y., VON ZASTROW, M. & KROGAN, N. J. 2017. An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell,* 169**,** 350-360. e12.

LOPEZ-BORGES, S. & LAZO, P. A. 2000. The human vaccinia-related kinase 1 (VRK1) phosphorylates threonine-18 within the mdm-2 binding site of the p53 tumour suppressor protein. *Oncogene,* 19**,** 3656-3664.

LOWARY, P. T. & WIDOM, J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning11Edited by T. Richmond. *Journal of Molecular Biology,* 276**,** 19-42.

LU, J. Y., LIN, Y. Y., BOEKE, J. D. & ZHU, H. 2013. Using functional proteome microarrays to study protein lysine acetylation. *Methods Mol Biol,* 981**,** 151-65.

LU, L., MILLIKIN, R. J., SOLNTSEV, S. K., ROLFS, Z., SCALF, M., SHORTREED, M. R. & SMITH, L. M. 2018. Identification of MS-cleavable and noncleavable chemically cross-linked peptides with MetaMorpheus. *Journal of proteome research,* 17**,** 2370-2376.

LUNDE, B. M., MOORE, C. & VARANI, G. 2007. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol,* 8**,** 479-90.

MACBEATH, G. & SCHREIBER, S. L. 2000. Printing proteins as microarrays for high-throughput function determination. *Science,* 289**,** 1760-3.

MAEKAWA, T., SAKURA, H., SUDO, T. & ISHII, S. 1989. Putative metal finger structure of the human immunodeficiency virus type 1 enhancer binding protein HIV-EP1. *Journal of Biological Chemistry,* 264**,** 14591-14593.

MARAZZI, I., HO, J. S. Y., KIM, J., MANICASSAMY, B., DEWELL, S., ALBRECHT, R. A., SEIBERT, C. W., SCHAEFER, U., JEFFREY, K. L., PRINJHA, R. K., LEE, K., GARCÍA-SASTRE, A., ROEDER, R. G. & TARAKHOVSKY, A. 2012. Suppression of the antiviral response by an influenza histone mimic. *Nature,* 483**,** 428-433.

MAURER-STROH, S., DICKENS, N. J., HUGHES-DAVIES, L., KOUZARIDES, T., EISENHABER, F. & PONTING, C. P. 2003. The Tudor domain 'Royal Family': tudor, plant agenet, chromo, PWWP and MBT domains. *Trends in biochemical sciences,* 28**,** 69-74.

MEYER, H., BUG, M. & BREMER, S. 2012. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biology,* 14**,** 117-123.

MINSKY, N. & OREN, M. 2004. The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression. *Molecular cell,* 16**,** 631-639.

MINSKY, N., SHEMA, E., FIELD, Y., SCHUSTER, M., SEGAL, E. & OREN, M. 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature cell biology,* 10**,** 483-488.

MIOTTO, B. & STRUHL, K. 2011. JNK1 phosphorylation of Cdt1 inhibits recruitment of HBO1 histone acetylase and blocks replication licensing in response to stress. *Molecular cell,* 44**,** 62-71.

MOODY, R. R., LO, M.-C., MEAGHER, J. L., LIN, C.-C., STEVERS, N. O., TINSLEY, S. L., JUNG, I., MATVEKAS, A., STUCKEY, J. A. & SUN, D. 2018. Probing the

interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. *The Journal of biological chemistry,* 293**,** 2125-2136.

MOYAL, L., LERENTHAL, Y., GANA-WEISZ, M., MASS, G., SO, S., WANG, S.-Y., EPPINK, B., CHUNG, Y. M., SHALEV, G. & SHEMA, E. 2011. Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Molecular cell,* 41**,** 529-542.

MULLALLY, J. E., CHERNOVA, T. & WILKINSON, K. D. 2006. Doa1 is a Cdc48 adapter that possesses a novel ubiquitin binding domain. *Molecular and cellular biology,* 26**,** 822-830.

NAKAMURA, K., KATO, A., KOBAYASHI, J., YANAGIHARA, H., SAKAMOTO, S., OLIVEIRA, D. V., SHIMADA, M., TAUCHI, H., SUZUKI, H. & TASHIRO, S. 2011. Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Molecular cell,* 41**,** 515-528.

NG, H. H., ROBERT, F., YOUNG, R. A. & STRUHL, K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell,* 11**,** 709-719.

NICASSIO, F., CORRADO, N., VISSERS, J. H., ARECES, L. B., BERGINK, S., MARTEIJN, J. A., GEVERTS, B., HOUTSMULLER, A. B., VERMEULEN, W. & DI FIORE, P. P. 2007. Human USP3 is a chromatin modifier required for S phase progression and genome stability. *Current biology,* 17**,** 1972-1977.

NICHOLS, R. J., WIEBE, M. S. & TRAKTMAN, P. 2006. The vaccinia-related kinases phosphorylate the N′ terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Molecular biology of the cell,* 17**,** 2451-2464.

NISHIOKA, K., RICE, J. C., SARMA, K., ERDJUMENT-BROMAGE, H., WERNER, J., WANG, Y., CHUIKOV, S., VALENZUELA, P., TEMPST, P. & STEWARD, R. 2002. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Molecular cell,* 9**,** 1201-1213.

OGISO, Y., SUGIURA, R., KAMO, T., YANAGIYA, S., LU, Y., OKAZAKI, K., SHUNTOH, H. & KUNO, T. 2004. Lub1 participates in ubiquitin homeostasis and stress response via maintenance of cellular ubiquitin contents in fission yeast. *Molecular and cellular biology,* 24**,** 2324-2331.

OHTAKE, F., SAEKI, Y., SAKAMOTO, K., OHTAKE, K., NISHIKAWA, H., TSUCHIYA, H., OHTA, T., TANAKA, K. & KANNO, J. 2015. Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO reports,* 16**,** 192-201.

OLSON, M. F. & SAHAI, E. 2009. The actin cytoskeleton in cancer cell motility. *Clinical & experimental metastasis,* 26**,** 273.

OOI, S. K., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S.-P. & ALLIS, C. D. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-717.

ORG, T., CHIGNOLA, F., HETÉNYI, C., GAETANI, M., REBANE, A., LIIV, I., MARAN, U., MOLLICA, L., BOTTOMLEY, M. J. & MUSCO, G. 2008. The autoimmune regulator PHD finger binds to non‐methylated histone H3K4 to activate gene expression. *EMBO reports,* 9**,** 370-376.

PAPADOPOULOS, C., KIRCHNER, P., BUG, M., GRUM, D., KOERVER, L., SCHULZE, N., POEHLER, R., DRESSLER, A., FENGLER, S. & ARHZAOUY, K. 2017.

VCP/p97 cooperates with YOD 1, UBXD 1 and PLAA to drive clearance of ruptured lysosomes by autophagy. *The EMBO journal,* 36**,** 135-150.

PASHKOVA, N., GAKHAR, L., WINISTORFER, S. C., YU, L., RAMASWAMY, S. & PIPER, R. C. 2010. WD40 repeat propellers define a ubiquitin-binding domain that regulates turnover of F box proteins. *Molecular cell,* 40**,** 433-443.

PAVRI, R., ZHU, B., LI, G., TROJER, P., MANDAL, S., SHILATIFARD, A. & REINBERG, D. 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell,* 125**,** 703-717.

PHIZICKY, E. M. & FIELDS, S. 1995. Protein-protein interactions: methods for detection and analysis. *Microbiological reviews,* 59**,** 94-123.

POKHOLOK, D. K., HARBISON, C. T., LEVINE, S., COLE, M., HANNETT, N. M., LEE, T. I., BELL, G. W., WALKER, K., ROLFE, P. A. & HERBOLSHEIMER, E. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell,* 122**,** 517-527.

POVLSEN, L. K., BELI, P., WAGNER, S. A., POULSEN, S. L., SYLVESTERSEN, K. B., POULSEN, J. W., NIELSEN, M. L., BEKKER-JENSEN, S., MAILAND, N. & CHOUDHARY, C. 2012. Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nature cell biology,* 14**,** 1089-1098.

PUIG, O., CASPARY, F., RIGAUT, G., RUTZ, B., BOUVERET, E., BRAGADO-NILSSON, E., WILM, M. & SÉRAPHIN, B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods,* 24**,** 218-229.

QIN, S., LIU, Y., TEMPEL, W., ERAM, M. S., BIAN, C., LIU, K., SENISTERRA, G., CROMBET, L., VEDADI, M. & MIN, J. 2014. Structural basis for histone

mimicry and hijacking of host proteins by influenza virus protein NS1. *Nature Communications,* 5**,** 3952.

RACINE, A., PAGÉ, V., NAGY, S., GRABOWSKI, D. & TANNY, J. C. 2012. Histone H2B ubiquitylation promotes activity of the intact Set1 histone methyltransferase complex in fission yeast. *Journal of Biological Chemistry,* 287**,** 19040-19047.

RAMACHANDRAN, N., HAINSWORTH, E., BHULLAR, B., EISENSTEIN, S., ROSEN, B., LAU, A. Y., WALTER, J. C. & LABAER, J. 2004. Self-assembling protein microarrays. *Science,* 305**,** 86-90.

RAMACHANDRAN, S., HADDAD, D., LI, C., LE, M. X., LING, A. K., SO, C. C., NEPAL, R. M., GOMMERMAN, J. L., YU, K. & KETELA, T. 2016. The SAGA deubiquitination module promotes DNA repair and class switch recombination through ATM and DNAPK-mediated γH2AX formation. *Cell reports,* 15**,** 1554-1565.

RAMÓN-MAIQUES, S., KUO, A. J., CARNEY, D., MATTHEWS, A. G., OETTINGER, M. A., GOZANI, O. & YANG, W. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proceedings of the National Academy of Sciences,* 104**,** 18993-18998.

RAO, V. S., SRINIVAS, K., SUJINI, G. & KUMAR, G. 2014. Protein-protein interaction detection: methods and analysis. *International journal of proteomics,* 2014.

REDMAN, K. L. & RECHSTEINER, M. 1989. Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature,* 338**,** 438-440.

RIBARDO, D. A., CROWE, S. E., KUHL, K. R., PETERSON, J. W. & CHOPRA, A. K. 2001. Prostaglandin levels in stimulated macrophages are controlled by phospholipase A2-activating protein and by activation of phospholipase C and D. *Journal of Biological Chemistry,* 276**,** 5467-5475.

RICHARDS, A. L., ECKHARDT, M. & KROGAN, N. J. 2021. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Molecular Systems Biology,* 17**,** e8792.

RICHARDSON, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem,* 34**,** 167-339.

RIGAUT, G., SHEVCHENKO, A., RUTZ, B., WILM, M., MANN, M. & SÉRAPHIN, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nature biotechnology,* 17**,** 1030-1032.

RUSSELL, N. S. & WILKINSON, K. D. 2004. Identification of a novel 29-linked polyubiquitin binding protein, Ufd3, using polyubiquitin chain analogues. *Biochemistry,* 43**,** 4844-4854.

RUTHENBURG, A. J., ALLIS, C. D. & WYSOCKA, J. 2007a. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell,* 25**,** 15-30.

RUTHENBURG, ALEXANDER J., LI, H., MILNE, THOMAS A., DEWELL, S., MCGINTY, ROBERT K., YUEN, M., UEBERHEIDE, B., DOU, Y., MUIR, TOM W., PATEL, DINSHAW J. & ALLIS, C. D. 2011. Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell,* 145**,** 692-706.

RUTHENBURG, A. J., LI, H., PATEL, D. J. & ALLIS, C. D. 2007b. Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews Molecular cell biology,* 8**,** 983-994.

SAHTOE, D. D., VAN DIJK, W. J., EKKEBUS, R., OVAA, H. & SIXMA, T. K. 2016. BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nature communications,* 7**,** 1-13.

SAMPATH, S. C., MARAZZI, I., YAP, K. L., SAMPATH, S. C., KRUTCHINSKY, A. N., MECKLENBRÄUKER, I., VIALE, A., RUDENSKY, E., ZHOU, M.-M. & CHAIT, B. T. 2007. Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Molecular cell,* 27**,** 596-608.

SANJANA, N. E., SHALEM, O. & ZHANG, F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods,* 11**,** 783-784.

SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. T., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-411.

SASTRY, M., ZHOU, W. & BANEYX, F. 2009. Integrity of N‐and C‐termini is important for E. coli Hsp31 chaperone activity. *Protein Science,* 18**,** 1439-1447.

SCHNEIDER, R., BANNISTER, A. J., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature cell biology,* 6**,** 73-77.

SCHÜBELER, D., MACALPINE, D. M., SCALZO, D., WIRBELAUER, C., KOOPERBERG, C., VAN LEEUWEN, F., GOTTSCHLING, D. E., O'NEILL, L. P., TURNER, B. M. & DELROW, J. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & development,* 18**,** 1263-1271.

SCHWER, B., ECKERSDORFF, M., LI, Y., SILVA, J. C., FERMIN, D., KURTEV, M. V., GIALLOURAKIS, C., COMB, M. J., ALT, F. W. & LOMBARD, D. B. 2009. Calorie restriction alters mitochondrial protein acetylation. *Aging cell,* 8**,** 604-606.

SCOTT, D., OLDHAM, N. J., STRACHAN, J., SEARLE, M. S. & LAYFIELD, R. 2015. Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics,* 15**,** 844-61.

SEIBERLICH, V., GOLDBAUM, O., ZHUKAREVA, V. & RICHTER-LANDSBERG, C. 2012. The small molecule inhibitor PR-619 of deubiquitinating enzymes affects the microtubule network and causes protein aggregate formation in neural cells: implications for neurodegenerative diseases. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research,* 1823**,** 2057-2068.

SEVILLA, A., SANTOS, C. R., VEGA, F. M. & LAZO, P. A. 2004. Human vaccinia-related kinase 1 (VRK1) activates the ATF2 transcriptional activity by novel phosphorylation on Thr-73 and Ser-62 and cooperates with JNK. *Journal of Biological Chemistry,* 279**,** 27458-27465.

SHALEM, O., SANJANA, N. E., HARTENIAN, E., SHI, X., SCOTT, D. A., MIKKELSEN, T. S., HECKL, D., EBERT, B. L., ROOT, D. E., DOENCH, J. G. & ZHANG, F. 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science,* 343**,** 84-87.

SHERMAN, F., STEWART, J. W. & TSUNASAWA, S. 1985. Methionine or not methionine at the beginning of a protein. *Bioessays,* 3**,** 27-31.

SHI, X., HONG, T., WALTER, K. L., EWALT, M., MICHISHITA, E., HUNG, T., CARNEY, D., PENA, P., LAN, F., KAADIGE, M. R., LACOSTE, N., CAYROU, C., DAVRAZOU, F., SAHA, A., CAIRNS, B. R., AYER, D. E., KUTATELADZE, T. G., SHI, Y., COTE, J., CHUA, K. F. & GOZANI, O. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature,* 442**,** 96-9.

SHI, X., KACHIRSKAIA, I., WALTER, K. L., KUO, J. H., LAKE, A., DAVRAZOU, F., CHAN, S. M., MARTIN, D. G., FINGERMAN, I. M., BRIGGS, S. D., HOWE, L., UTZ,

P. J., KUTATELADZE, T. G., LUGOVSKOY, A. A., BEDFORD, M. T. & GOZANI, O. 2007. Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem,* 282**,** 2450-5.

SHIBA, Y., KATOH, Y., SHIBA, T., YOSHINO, K., TAKATSU, H., KOBAYASHI, H., SHIN, H. W., WAKATSUKI, S. & NAKAYAMA, K. 2004. GAT (GGA and Tom1) domain responsible for ubiquitin binding and ubiquitination. *J Biol Chem,* 279**,** 7105-11.

SHIH, S. C., PRAG, G., FRANCIS, S. A., SUTANTO, M. A., HURLEY, J. H. & HICKE, L. 2003. A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. *EMBO J,* 22**,** 1273-81.

SHILOH, Y., SHEMA, E., MOYAL, L. & OREN, M. 2011. RNF20–RNF40: A ubiquitin-driven link between gene expression and the DNA damage response. *FEBS letters,* 585**,** 2795-2802.

SIDOLI, S. & GARCIA, B. A. 2017. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert review of proteomics,* 14**,** 617-626.

SIGGERS, T. & GORDAN, R. 2014. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res,* 42**,** 2099-111.

SIMS, J. J. & COHEN, R. E. 2009. Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Mol Cell,* 33**,** 775-83.

SIMS, R. J., CHEN, C.-F., SANTOS-ROSA, H., KOUZARIDES, T., PATEL, S. S. & REINBERG, D. 2005. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry,* 280**,** 41789-41792.

SMITH, M. G., JONA, G., PTACEK, J., DEVGAN, G., ZHU, H., ZHU, X. & SNYDER, M. 2005. Global analysis of protein function using protein microarrays. *Mech Ageing Dev,* 126**,** 171-5.

SONG, J.-J. & KINGSTON, R. E. 2008. WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *Journal of Biological Chemistry,* 283**,** 35258-35264.

SONG, O.-K., WANG, X., WATERBORG, J. H. & STERNGLANZ, R. 2003. An Nα-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A. *Journal of Biological Chemistry,* 278**,** 38109-38112.

SONGYANG, Z., SHOELSON, S. E., CHAUDHURI, M., GISH, G., PAWSON, T., HASER, W. G., KING, F., ROBERTS, T., RATNOFSKY, S., LECHLEIDER, R. J. & ET AL. 1993. SH2 domains recognize specific phosphopeptide sequences. *Cell,* 72**,** 767-78.

STRAHL, B. D. & ALLIS, C. D. 2000a. The language of covalent histone modifications. *Nature,* 403**,** 41-45.

STRAHL, B. D. & ALLIS, C. D. 2000b. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

SVENSSON, E. C., HUGGINS, G. S., DARDIK, F. B., POLK, C. E. & LEIDEN, J. M. 2000. A Functionally Conserved N-terminal Domain of the Friend of GATA-2 (FOG-2) Protein Represses GATA4-Dependent Transcription. *Journal of Biological Chemistry,* 275**,** 20762-20769.

SWANEY, D. L., RODRÍGUEZ-MIAS, R. A. & VILLÉN, J. 2015. Phosphorylation of ubiquitin at Ser65 affects its polymerization, targets, and proteome-wide turnover. *EMBO reports,* 16**,** 1131-1144.

TAVERNA, S. D., ILIN, S., ROGERS, R. S., TANNY, J. C., LAVENDER, H., LI, H., BAKER, L., BOYLE, J., BLAIR, L. P. & CHAIT, B. T. 2006. Yng1 PHD finger binding to

H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Molecular cell,* 24**,** 785-796.

TAVERNA, S. D., LI, H., RUTHENBURG, A. J., ALLIS, C. D. & PATEL, D. J. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology,* 14**,** 1025-1040.

TELONI, F. & ALTMEYER, M. 2016. Readers of poly(ADP-ribose): designed to be fit for purpose. *Nucleic Acids Res,* 44**,** 993-1006.

TOMLINSON, V., GUDMUNDSDOTTIR, K., LUONG, P., LEUNG, K., KNEBEL, A. & BASU, S. 2010. JNK phosphorylates Yes-associated protein (YAP) to regulate apoptosis. *Cell death & disease,* 1**,** e29-e29.

TROJER, P., LI, G., SIMS, R. J., 3RD, VAQUERO, A., KALAKONDA, N., BOCCUNI, P., LEE, D., ERDJUMENT-BROMAGE, H., TEMPST, P., NIMER, S. D., WANG, Y. H. & REINBERG, D. 2007. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell,* 129**,** 915-28.

TWEEDIE-CULLEN, R. Y., RECK, J. M. & MANSUY, I. M. 2009. Comprehensive mapping of post-translational modifications on synaptic, nuclear, and histone proteins in the adult mouse brain. *Journal of proteome research,* 8**,** 4966-4982.

VALBUENA, A., LÓPEZ-SÁNCHEZ, I. & LAZO, P. A. 2008. Human VRK1 is an early response gene and its loss causes a block in cell cycle progression. *PloS one,* 3**,** e1642.

VAN INGEN, H., VAN SCHAIK, F. M., WIENK, H., BALLERING, J., REHMANN, H., DECHESNE, A. C., KRUIJZER, J. A., LISKAMP, R. M., TIMMERS, H. T. M. & BOELENS, R. 2008. Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure,* 16**,** 1245-1256.

VASSILEV, A. P., RASMUSSEN, H. H., CHRISTENSEN, E. I., NIELSEN, S. & CELIS, J. E. 1995. The levels of ubiquitinated histone H2A are highly upregulated in transformed human cells: partial colocalization of uH2A clusters and PCNA/cyclin foci in a fraction of cells in S-phase. *Journal of Cell Science,* 108**,** 1205-1215.

VAUGHAN, R. M., KUPAI, A., FOLEY, C. A., SAGUM, C. A., TIBBEN, B. M., EDEN, H. E., TIEDEMANN, R. L., BERRYHILL, C. A., PATEL, V. & SHAW, K. M. 2020. The histone and non-histone methyllysine reader activities of the UHRF1 tandem Tudor domain are dispensable for the propagation of aberrant DNA methylation patterning in cancer cells. *Epigenetics & chromatin,* 13**,** 1-15.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. P., VAN SCHAIK, F. M., VARIER, R. A., BALTISSEN, M. P., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. M. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell,* 131**,** 58-69.

WALSER, F., MULDER, M. P., BRAGANTINI, B., BURGER, S., GUBSER, T., GATTI, M., BOTUYAN, M. V., VILLA, A., ALTMEYER, M. & NERI, D. 2020. Ubiquitin Phosphorylation at Thr12 Modulates the DNA Damage Response. *Molecular Cell,* 80**,** 423-436. e9.

WANG, G. G., SONG, J., WANG, Z., DORMANN, H. L., CASADIO, F., LI, H., LUO, J.-L., PATEL, D. J. & ALLIS, C. D. 2009a. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature,* 459**,** 847-851.

WANG, H., CAO, R., XIA, L., ERDJUMENT-BROMAGE, H., BORCHERS, C., TEMPST, P. & ZHANG, Y. 2001. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Molecular cell,* 8**,** 1207-1217.

WANG, Y., REDDY, B., THOMPSON, J., WANG, H., NOMA, K., YATES, J. R., 3RD & JIA, S. 2009b. Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol Cell,* 33**,** 428-37.

WAUER, T., SWATEK, K. N., WAGSTAFF, J. L., GLADKOVA, C., PRUNEDA, J. N., MICHEL, M. A., GERSCH, M., JOHNSON, C. M., FREUND, S. M. & KOMANDER, D. 2015. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *The EMBO Journal,* 34**,** 307-325.

WEN, H., LI, J., SONG, T., LU, M., KAN, P.-Y., LEE, M. G., SHA, B. & SHI, X. 2010. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *Journal of Biological Chemistry,* 285**,** 9322-9326.

WEST, M. H. & BONNER, W. M. 1980. Histone 2B can be modified by the attachment of ubiquitin. *Nucleic acids research,* 8**,** 4671-4680.

WETLAUFER, D. B. 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A,* 70**,** 697-701.

WILKINSON, C. R., SEEGER, M., HARTMANN-PETERSEN, R., STONE, M., WALLACE, M., SEMPLE, C. & GORDON, C. 2001. Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat Cell Biol,* 3**,** 939-43.

WINGFIELD, P. T. 2017a. N-Terminal Methionine Processing. *Current Protocols in Protein Science,* 88**,** 6.14.1-6.14.3.

WINGFIELD, P. T. 2017b. N‐Terminal methionine processing. *Current protocols in protein science,* 88**,** 6.14. 1-6.14. 3.

WU, H., MATHIOUDAKIS, N., DIAGOURAGA, B., DONG, A., DOMBROVSKI, L., BAUDAT, F., CUSACK, S., DE MASSY, B. & KADLEC, J. 2013a. Molecular basis for

the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell reports,* 5**,** 13-20.

WU, M., WANG, L., LI, Q., LI, J., QIN, J. & WONG, J. 2013b. The MTA family proteins as novel histone H3 binding proteins. *Cell & bioscience,* 3**,** 1-1.

WU, X., LI, L. & JIANG, H. 2016. Doa1 targets ubiquitinated substrates for mitochondria-associated degradation. *Journal of Cell Biology,* 213**,** 49-63.

WU, Z., CHENG, Z., SUN, M., WAN, X., LIU, P., HE, T., TAN, M. & ZHAO, Y. 2015. A Chemical Proteomics Approach for Global Analysis of Lysine Monomethylome Profiling. *Molecular & Cellular Proteomics,* 14**,** 329-339.

WYSOCKA, J., SWIGUT, T., MILNE, T. A., DOU, Y., ZHANG, X., BURLINGAME, A. L., ROEDER, R. G., BRIVANLOU, A. H. & ALLIS, C. D. 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell,* 121**,** 859-72.

WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T., BADENHORST, P., WU, C. & ALLIS, C. D. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature,* 442**,** 86-90.

XU, J., BAUER, D. E., KERENYI, M. A., VO, T. D., HOU, S., HSU, Y.-J., YAO, H., TROWBRIDGE, J. J., MANDEL, G. & ORKIN, S. H. 2013. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proceedings of the National Academy of Sciences,* 110**,** 6518-6523.

YAMAKAMI, M., YOSHIMORI, T. & YOKOSAWA, H. 2003. Tom1, a VHS domain-containing protein, interacts with tollip, ubiquitin, and clathrin. *J Biol Chem,* 278**,** 52865-72.

YANG, N., WANG, W., WANG, Y., WANG, M., ZHAO, Q., RAO, Z., ZHU, B. & XU, R. M. 2012. Distinct mode of methylated lysine-4 of histone H3 recognition by

tandem tudor-like domains of Spindlin1. *Proc Natl Acad Sci U S A,* 109**,** 17954-9.

YANG, Y., LU, Y., ESPEJO, A., WU, J., XU, W., LIANG, S. & BEDFORD, M. T. 2010. TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell,* 40**,** 1016-23.

YAP, K. L. & ZHOU, M.-M. 2011. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry,* 50**,** 1966-1980.

YAU, R. & RAPE, M. 2016. The increasing complexity of the ubiquitin code. *Nat Cell Biol,* 18**,** 579-86.

YU, C. & HUANG, L. 2018. Cross-linking mass spectrometry (XL-MS): An emerging technology for interactomics and structural biology. *Analytical chemistry,* 90**,** 144.

YU, Y., ZHENG, Q., ERRAMILLI, S. K., PAN, M., PARK, S., XIE, Y., LI, J., FEI, J., KOSSIAKOFF, A. A. & LIU, L. 2020. Interaction Landscape of K29-Linked Ubiquitin Signaling Revealed by a Linkage-Specific Synthetic Antigen-Binding Fragment. *bioRxiv*.

ZEGERMAN, P., CANAS, B., PAPPIN, D. & KOUZARIDES, T. 2002. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *Journal of Biological Chemistry,* 277**,** 11621-11624.

ZENG, L., ZHANG, Q., LI, S., PLOTNIKOV, A. N., WALSH, M. J. & ZHOU, M.-M. 2010. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature,* 466**,** 258-262.

ZHANG, X., PENG, D., XI, Y., YUAN, C., SAGUM, C. A., KLEIN, B. J., TANAKA, K., WEN, H., KUTATELADZE, T. G., LI, W., BEDFORD, M. T. & SHI, X. 2016. G9a-

mediated methylation of ERalpha links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression. *Nat Commun,* 7**,** 10810.

ZHANG, X.-Y., VARTHI, M., SYKES, S. M., PHILLIPS, C., WARZECHA, C., ZHU, W., WYCE, A., THORNE, A. W., BERGER, S. L. & MCMAHON, S. B. 2008. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. *Molecular cell,* 29**,** 102-111.

ZHANG, Y., SHAN, C.-M., WANG, J., BAO, K., TONG, L. & JIA, S. 2017a. Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Scientific Reports,* 7**,** 43906.

ZHANG, Z., JONES, A., JOO, H.-Y., ZHOU, D., CAO, Y., CHEN, S., ERDJUMENT-BROMAGE, H., RENFROW, M., HE, H. & TEMPST, P. 2013. USP49 deubiquitinates histone H2B and regulates cotranscriptional pre-mRNA splicing. *Genes & development,* 27**,** 1581-1595.

ZHANG, Z., JONES, A. E., WU, W., KIM, J., KANG, Y., BI, X., GU, Y., POPOV, I. K., RENFROW, M. B. & VASSYLYEVA, M. N. 2017b. Role of remodeling and spacing factor 1 in histone H2A ubiquitination-mediated gene silencing. *Proceedings of the National Academy of Sciences,* 114**,** E7949-E7958.

ZHAO, G., LI, G., SCHINDELIN, H. & LENNARZ, W. J. 2009. An Armadillo motif in Ufd3 interacts with Cdc48 and is involved in ubiquitin homeostasis and protein degradation. *Proceedings of the National Academy of Sciences,* 106**,** 16197-16202.

ZHAO, S., YANG, M., ZHOU, W., ZHANG, B., CHENG, Z., HUANG, J., ZHANG, M., WANG, Z., WANG, R., CHEN, Z., ZHU, J. & LI, H. 2017. Kinetic and high-throughput profiling of epigenetic interactions by 3D-carbene chip-based surface

plasmon resonance imaging technology. *Proc Natl Acad Sci U S A,* 114**,** E7245-E7254.

ZHAO, Y., LANG, G., ITO, S., BONNET, J., METZGER, E., SAWATSUBASHI, S., SUZUKI, E., LE GUEZENNEC, X., STUNNENBERG, H. G. & KRASNOV, A. 2008. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Molecular cell,* 29**,** 92-101.

ZHOU, W., ZHU, P., WANG, J., PASCUAL, G., OHGI, K. A., LOZACH, J., GLASS, C. K. & ROSENFELD, M. G. 2008. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Molecular cell,* 29**,** 69-80.

ZHU, B., ZHENG, Y., PHAM, A.-D., MANDAL, S. S., ERDJUMENT-BROMAGE, H., TEMPST, P. & REINBERG, D. 2005. Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Molecular cell,* 20**,** 601-611.

ZHU, H., BILGIN, M., BANGHAM, R., HALL, D., CASAMAYOR, A., BERTONE, P., LAN, N., JANSEN, R., BIDLINGMAIER, S., HOUFEK, T., MITCHELL, T., MILLER, P., DEAN, R. A., GERSTEIN, M. & SNYDER, M. 2001. Global analysis of protein activities using proteome chips. *Science,* 293**,** 2101-5.

VITA

Jianji Chen was born in Baotou, China. He received his Bachelor's degree in Environmental Science at Dalian University of Technology (Dalian, China) in 2013. He then entered the University of Minnesota (UMN) at Twin Cities (Minneapolis, MN) and received his Master's degree in Environmental Health Science-Toxicology in 2015. After graduation, he became a junior scientist developing quantitative proteomics methods using high-resolution mass spectrometers at the Center for Mass Spectrometry and Proteomics at UMN. In August 2016, he entered the University of Texas M.D. Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences and joined Dr. Mark Bedford's group in 2017.

Permanent address:

9303 McKnight Loop,

Austn, TX, 78717